CITY UNIVERSITY LONDON

Assessing Early Sociocognitive and Language Skills in Young Saudi Children

Aseel AlKadhi

Thesis submitted for the degree of Doctor of Philosophy

City University London

Department of Language and Communication Science

May 2015

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisors Professor Shula Chiat and Professor Penny Roy for their unlimited support and invaluable guidance. They were always available for my questions, gave generously of their time and provided me with constructive feedback on various drafts of this thesis. Their enthusiasm, encouragement and faith in me have been extremely helpful.

I must also thank my friend Ashwag Wallan who has been extremely supportive during the course of this project in so many ways. I am particularly grateful for her help in the development and adaptation of parental questionnaires and in doing the inter-reliability checks and for the many lively discussions we had throughout.

My thanks are also due to Nora Alsudairy and Ghaida Alsugair for their help in adapting the parental questionnaire to Arabic, to the children and their families who participated in the study, and to the nurseries who allowed me to carry out the study.

A special thank you to my daughter Deema, who helped me draw the pictures for the vocabulary test.

I am very grateful to my parents, my mother-in-law and my brothers who have been a constant source of love, support and encouragement, to my children for their patience and understanding during the many times I had to be away to work. They have been my greatest source of motivation to finish this project.

Finally, words cannot express how much I appreciate the love and tremendous support offered by my husband Abdulaziz. I thank him for being my best friend and an amazing father to our children. I could not have done this without him.

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

ABSTRACT

Children with early language delay form a heterogeneous group. Although a significant number will catch up and develop language in the normal range, some will continue to have difficulties with language. Predicting the outcome for these children represents a challenging task for clinicians.

It has been suggested that the assessment of sociocognitive skills contributes distinctively to the prediction of persistence of language and communication difficulties and the nature of these difficulties. In the absence of standardized assessments in Saudi Arabia for children with early language delay, this study aimed to take a first step to filling this gap by developing a battery of early sociocognitive and language measures. The battery consisted of six measures assessing sociocognitive and language skills using direct and indirect methods, some existing and some newly developed or adapted for this project. Sociocognitive measures were the Early Sociocognitive Battery (ESB; Chiat & Roy, 2006b), together with a new Motor Imitation test (MI) and Sociocognitive Questionnaire (SCogQ); language measures included the Sentence Repetition test (Wallan, Chiat, & Roy, 2011), a new Arabic research adaptation of the Language Use Inventory (O'Neill, 2009), and a preschool adapted version of the Arabic Picture Vocabulary Test (Shaalan, 2010). Since this project was performed in a very different language culture and included a wider range of sociocognitive and language measures than most previous studies, a second aim was to investigate relations between the different sociocognitive and language skills.

The battery was administered to 161 Saudi children between the ages of 2;0-3;5 years, divided into three six-months age groups and almost equally divided into boys and girls.

Addressing the first aim of this study, results showed that all the measures with the exception of the SCogQ were reliable, valid, and age sensitive. These findings suggest that the measures are fit for purpose and have the potential to identify children with early language delay. Parental concern matched children's performance on direct and indirect measures of language for the majority of children.

Turning to the second aim of the study, regressional analyses using the three language assessments as outcome measures showed that the ESB and MI were important predictors of pragmatic language and receptive vocabulary when other measures had been taken into account.

It is concluded that the substantial set of data that this study has produced on the wide-ranging battery of assessments can serve as a reference for clinical comparison and as a foundation for standardization with a fully representative sample of young Saudi children. These measures not only enable the formal identification of a delay in Saudi preschoolers but are also informative about strengths and difficulties and can guide intervention. The results add to current understanding of the role sociocognitive skills play in language development, and provide the foundation for longitudinal research investigating relations to longer term outcomes.

ABBREVIATIONS

APS-RVT	Arabic Preschool-Receptive Vocabulary Test
ARA-LUI	Arabic Research Adaptation of the LUI
ASD	Autism spectrum disorders
BCa	Bias corrected and accelerated bootstrapping
CA	Chronological age
CDI	Communicative Development Inventories
CELF	Clinical Evaluation of Language Fundamentals
CHAT	Checklist for Autism in Toddlers
CI	Confidence intervals
CSBS-DP	Communication and Symbolic Behaviour Scale-Developmental Profile
ESB	Early Sociocognitive Battery
ESCS	Early Social and Communication Scale
ICC	Intraclass correlation coefficient
IQ	Intelligent quotient
LDS	Language Development Survey
LR+	Positive likelihood ratio
LR-	Negative likelihood ratio
LUI	Language Use Inventory
MI	Motor Imitation
MIS	Motor Imitation Scale
MLU	Mean length of utterance
NPV	Negative predicative value
NVMA	Nonverbal mental age
PLS	Preschool Language Scale
PPV	Positive predictive value
SCogQ	Sociocognitive Questionnaire
SES	Socio-economic status
SLI	Specific language impairment
SR	Sentence Repetition
SR Grammatical	Sentence repetition grammatical morpheme score
SR Lexical	Sentence repetition lexical morpheme score

ToM	Theory of Mind
TSA	Total sentence accuracy score

CHAPTER 1

INTRODUCTION

Toddlers with early language delay represent 10% to 17.5% of otherwise typically developing children (Horwitz et al., 2003; Rescorla, 1989). It is well recognized that early identification and intervention services may enable children with language and communication difficulties to reach their potential and in some cases prevent secondary complications (Broomfield & Dodd, 2011; Girolametto, Wiigs, Smyth, Weitzman, & Pearce, 2001). From a biological perspective, researchers have stressed that the greatest impact of language stimulation and intervention is during early childhood while the brain is still plastic (Paul & Roth, 2011). Children with language problems that are secondary to syndromes or established medical conditions are generally identified at birth and enrolled in an early intervention services. However, in the case of children who present with delayed language in the absence of a diagnosed physical or developmental disability, the decision of whether to intervene or not is more difficult given the variability of early language (Fenson et al., 2000; Gatt, Grech, & Dodd, 2013) and the significant number of late talkers who show spontaneous recovery (Dale, Price, Bishop, & Plomin, 2003). Ideally from a clinical perspective, services need to separate out children with transient language difficulties, also referred to as "*late bloomers"* (Rescorla, 1989) and identify and target children with persistent problems that do not spontaneously recover.

In the West, there is a substantial body of research following up late talking children in order to identify predictors of outcome. Researchers have indicated that relying only on language measures provides poor predictability in discriminating children with persistent difficulties from children with transient difficulties (Dale, Price, Bishop, & Plomin, 2003). It has been suggested that the assessment of early nonverbal sociocognitive skills will contribute distinctively to predicting subsequent language and communication development (Chiat & Roy, 2008; Eadie et al., 2013; Wetherby, Allen, Cleary, Kublin, & Goldstein, 2002). Chiat and Roy (2006a) have developed the Early Sociocognitive Battery (ESB) which provides information about children's joint attention, social responsiveness, and symbolic comprehension. They hypothesised that these early sociocognitive skills serve as predictors of the likelihood and nature of longer term language and communication difficulties in preschoolers (Chiat & Roy, 2008, 2013). Nonverbal

imitation is another measure of sociocognition that has been associated with language and communication (Carpenter, Nagell, & Tomasello, 1998).

These assessments of early sociocognition are primarily nonverbal, and as such lend themselves to cross-linguistic application. This is particularly useful in a language community where there are very few, if any, developmental norms for language as is the case in Saudi Arabic. Results of a previous study with a small number of typical Saudi preschool children showed that their scores on the ESB were very similar to the scores of British children of the same age (Alkadhi, 2010), a finding that suggests the appropriateness of this measure cross-culturally.

In the absence of standardized assessment tools for toddlers with language and communication difficulties in Saudi Arabia, the first aim of this study was to develop a battery to assess early language and nonverbal sociocognitive skills using measures that have been reported to be most informative of language outcome. The development of this battery will include establishing the reliability, validity and age sensitivity of the newly developed measures. Data collected on these measures will contribute to the establishment of norms which could guide clinical diagnosis and intervention with young Saudi children.

A second aim was to examine the associations and dissociations between the sociocognitive and language measures. Most of the published research on relations between early language and sociocognitive skills has been in Western cultures with English-speaking children and few have looked at the relative contribution of different skills of sociocognition to different language skills. Investigating these associations in Arabic children will provide new evidence on the informativeness of the different assessments of language and sociocognition cross-culturally, contributing to the theoretical understanding of the nature of the relationship between language skills and sociocognitive skills. In addition, it will provide an opportunity to advance our knowledge of the early predictors to children's early language and communication development.

Outline of the thesis

This chapter will provide an overview of studies with early language delay aged 16-48 months. This section draws on studies carried out in UK, US, Scandinavia, Canada and Australia. To date, research on early language delay in Saudi has been constrained by the dearth of language assessments in Saudi and lack of developmental norms. Behaviors that suggest an increased risk for language impairment will be identified and the current evidence on predictors will be summarized.

Chapter 2 provides an overview of studies examining relations between nonverbal sociocognitive skills assessed in the ESB and language.

Chapter 3 provides an overview of the relations between nonverbal imitation and language across different types of motor imitation and asks whether there are differences in performance of clinical groups-more specifically children with ASD or children with language delay-on motor imitation tasks.

Chapter 4 describes in detail how the measures used in this study were developed or adapted from existing measures.

Chapter 5 describes the methodology of the study including recruitment procedures, participants' characteristics and the administration and scoring procedure for the different tasks.

Results are presented in *Chapter 6* and an interpretation and discussion of the results are provided in *Chapter 7* as well as limitations and implications of the study.

1.1 Early Language Delay

1.1.1 Terms and criteria

Different terms have been used by researchers to early language delay such as specific expressive vocabulary delay (Rescorla & Schwartz, 1990), early language delay (Dale et al., 2003) and most commonly late talkers. These terms and more specifically the term *late talkers* have been used to refer to children ranging in age between 18-30 months (Paul & Roth, 2011) or 18-35 months (Rescorla, 2011) who are slow at producing their first words or word combinations in the absence of any other diagnosed disability such as hearing

impairment, emotional problems, or cognitive deficits. They may or may not have age appropriate receptive language abilities. Criteria used to identify late-talking vary across studies. Among the criteria used are: vocabulary size at or below the 10th percentile for age group on a parent report of expressive vocabulary (Dale et al., 2003; Henrichs et al., 2011), less than 50 words at 24-31 months and at least 6 months below chronological age (CA) on a standardized measure of expressive language (Rescorla, 1997); or lack of word combinations at 2 years (Poll & Miller, 2013).

Although a significant proportion (55-60%) of children who have been identified as late talkers will catch up after a slow start "*late bloomers*" and exhibit age appropriate language skills by kindergarten (Dale et al., 2003), early language delay might be an indication of specific language impairment (SLI) or autism (Buschmann et al., 2008). For these children early intervention is desirable. As a consequence, several researchers have followed up late talkers to identify predictors that differentiate late talkers into those who will catch up and those who will not. Investigators have employed different methodological approaches to address this issue. In one approach, researchers have recruited large numbers of children from the general population. The samples in these studies were demographically heterogeneous and largely resembled the population. In most cases, but not all, only children's language ability was measured, and parental reports were used. Thus, although by definition late talkers are children who show delay in expressive language in the absence of other problems, it is not possible to rule out that some of the children included in the late talkers' groups in these studies have additional problems such as low IQ. Studies using this approach will be referred to as large scale studies in the next section. In the other approach, only children who presented with language problems in the absence of additional problems were included. Those children were mainly identified through direct testing and were in most cases compared to a group of typically developing children. These studies will be referred to as small scale studies. The sociodemographic characteristics of the samples in small scale studies were sometimes biased towards low or high socioeconomic status (SES) depending on the recruitment procedure. On the other hand, large scale studies generally included children from varied SES, although it was not reported in all studies how the sample characteristics compared with the general population. In addition, due to differences in recruitment procedures and inclusion criteria, it is expected that large scale studies will include children with a larger range of language abilities compared to small scale studies.

Examining findings and drawing conclusions from both large and small scale studies and how they linkup will help to improve our understanding of language development and identify measures that are most predictive of outcome in children with concerns about their language. This will include longitudinal studies that examined outcome of children identified with language delay between the ages of 16 months to 4 years, thus, older than the children who have been commonly referred to as late talkers. However they are still within the age range when a significant number might show spontaneous recovery and clinicians might be uncertain whether to intervene or not. Children who present with language delay at 5 years are likely to continue to have language problems and intervention is almost always indicated (Rescorla, Dahlsgaard, & Roberts, 2000). Findings of large and small scale studies are discussed in the following sections.

A recent study (Pesco & O'Neill, 2012) evaluated the predictiveness of a pragmatic language assessment. Since it differed in many aspects from both large and small scale studies such as sample size, age range and recruitment procedure it will be reviewed in a separate section.

1.1.2 Large scale studies

1.1.2.1 Introduction

A number of studies in the UK, US, Netherlands, Sweden, and Australia have followed late talkers to identify predictors of language delay. These studies are population based studies, in which children were generally recruited through national population registration or child health services and sample sizes mostly exceeded 1000, aiming to include children from varied socioeconomic backgrounds as pointed out in the previous section. However, population samples may differ. The inclusion criteria have varied across studies, with some studies not reporting any exclusion criteria (Henrichs et al., 2011); some excluding children with major medical disorders, known diagnosis of autism, and genetic syndromes based on parents' reports (Dale et al., 2003); and others using more stringent criteria and only including healthy children with average intellectual abilities based on standardized measures at follow up (Poll & Miller, 2013). Furthermore, some studies have included participants from different language backgrounds (Henrichs et al., 2011) although most have indicated that they only included participants whose first language was the main language of the country (Dale et al., 2003; Westerlund, Berglund, &

Eriksson, 2006). These differences in the exclusion criteria might affect estimates of children with language delay reported in different studies.

In all these studies, irrespective of the nature of the samples, the contribution of multiple environmental, familial and child factors was examined. In addition, the child's language and communication characteristics were investigated. The language outcome of the children in the reviewed studies will be presented and significant predictors will be reported. However, the focus will be in summarizing results relating to the child's language and communication abilities as predictors.

In determining the accuracy of skills measured to predict outcome at case level, different statistical analyses were used by different investigators. In most studies sensitivity, specificity, positive and negative predictive values were reported. Sensitivity refers to the proportion of children with language impairment who were correctly predicted by the tests. Specificity refers to the proportion of children with normal outcome who had been identified as such by the predictor (Straus, Glasziou, Richardson, & Haynes, 2011). It has been suggested that the minimum acceptable value of sensitivity and specificity is 80%. Values of 90% or more are considered an indication of good classification accuracy in diagnosing language impairment (Plante & Vance, 1994). While sensitivity and specificity are considered fixed properties of the test, positive and negative predictive values are dependent on the prevalence of the disorder in the population (Sackett, Haynes, Guyatt, & Tugwell, 1991). Positive predictive value (PPV) is the proportion of children who are positive on the predictor who do have language impairment at outcome and negative predictive value (NPV) is the proportion of children who are negative on the predictor and are normal at outcome. An additional statistic that is sometimes reported and is independent of base rates is the likelihood ratio. A positive likelihood ratio (LR+) refers to the probability that a child who scored positive on the test truly has the target disorder and a negative likelihood ratio (LR-) refers to probability of having the target disorder given a negative test result (Straus, Glasziou, Richardson, & Haynes, 2011). According to Straus and colleagues, values of LR+ above 10 are considered large, between 5 and 10 are moderate and between 2 and 4 are small. Since the studies have usually used different exclusion criteria for late talkers, use of predictive values to compare studies is problematic.

1.1.2.2 Findings of large scale studies

Dale and colleagues (2003) assessed 8386 twins in the UK at 2 years of age and then again at 3 and 4. At age 2, children's vocabulary, grammar, displaced reference (which included items that assess child's ability to talk about past and future), and nonverbal ability were assessed using parental reports. Children who scored below the 10% percentile on a parental report of expressive vocabulary at age two were identified as late talkers. Outcome language measures at 3 and 4 years of age were parental reports of vocabulary, grammar, and use of abstract language. Children who scored below the 15th percentile on at least 2 of the 3 language measures were identified as language impaired. Results showed that 56% of late talkers in this sample of typically developing children caught up by 3 years and 60% did so by 4. At 2 years of age, the majority (96%) of late talkers scored 0 on the grammar measure and were not combining words, hence no significant relations were found between grammar at 2 years and outcome at 3 and 4 years in the late talker group. On the other hand, significant relations were found between vocabulary, displaced reference and nonverbal abilities at 2 years of age and language outcome at 3 and 4 years. Logistic regression was used to predict language outcome. Within the late talker group, the model of predictors which included age 2 vocabulary, displaced reference and nonverbal abilities based on parental report demonstrated high specificity (80.5%) in predicting outcome at age 4. However, the sensitivity (44.6%) and positive and negative predictive values were low (61.1% and 67.7% respectively). Thus, classification of outcome based on language delay at 2 years of age failed to detect the majority of children with language impairment at age 4.

Similarly, in Australia, Reilly and colleagues (2010) examined the contribution of late talking status at 24 months in addition to a number of demographic, maternal and perinatal factors to language outcome at 4 years of age. They used the same criterion adopted in Dale and colleagues' study for classifying late talkers (below 10% percentile on parental report of vocabulary). Impairment at outcome was defined as more than 1.25 below the mean for the normative population on the receptive or expressive parts of the Australian adaptation of the Clinical Evaluation of Language Fundamentals-Preschool, Second Edition (CELF-P2; Wiig, Secord & Semel, 2006). Follow-up data was provided for 1596 children. Multivariate linear regression analysis was performed for a number of demographic, perinatal and maternal factors. Results showed that adding in late talking

status improved the prediction especially for the expressive language domain, though it only explained an additional 9.5% of the variance.

A similar picture emerged in the Netherlands. Henrichs and colleagues (2011) examined numerous biological, environmental and child factors as predictors of continuity of vocabulary skills in a population-based cohort study. Their sample of 3,759 children included those who spoke Dutch, English, Turkish and Arabic. However, unlike in the previous studies, late talking was determined by vocabulary scores at 18 months (using the same cutoff of scoring below the 10^{th} percentile for age). Initial assessment included parental reports of both receptive and expressive vocabulary. At 30 months, only expressive vocabulary was measured using parental reports. Again, children with scores below the 10^{th} percentile were considered delayed at outcome. Results showed that most of the children (71%) who were delayed at 18 months caught up at follow-up. Correlational analyses showed moderate correlations between expressive vocabulary at 18 and 30 months ($r = .34$, $p < .001$) and small correlations between receptive vocabulary at 18 months and expressive vocabulary at 30 months ($r = .19$, $p < .001$). In line with Reilly and colleagues' study, hierarchical linear regression showed that expressive vocabulary accounted for only 11% of the variance in outcome. Receptive vocabulary explained an additional 0.5% of the variance. In addition, results showed that predictiveness of outcome based on classification of late talking at 18 months was not strong, with low sensitivity (30%) and low positive predictive value (29%). Based on vocabulary scores at intake and follow-up, children were categorized into four groups: a normal group, a late bloomer group (delayed at intake but normal at outcome), a late onset delay group (normal at intake but delayed at outcome) and a persistent delay group (delayed at both intake and outcome). The independent value of the different factors to outcome was also examined using logistic regression. Results showed that relative to the group with normal language, children with receptive vocabulary delay at 18 months had a 9 times higher risk of being in the persistent vocabulary delay group and 4 times higher risk of being a late bloomer or developing a later vocabulary delay.

Likewise, in Sweden, Westerlund, Berglund and Eriksson (2006) examined the effectiveness of identification as a late talker at 18 months based on parental reports for predicting language delay at 3 years of age. A criterion of less than 8 words was used to define children as delayed at 18 months. Children's word production, comprehension and gesture use were assessed at intake through parental reports. Children's language outcome

at 3 years of age was measured by nurses on the basis of formalized observations of receptive and expressive language. Impairment at outcome was defined as inability to produce three-word sentences or inability to show comprehension of 3 out of 5 questions by responding verbally or pointing to pictures. Data at intake and follow-up was available for 891 children. Using logistic regression they found that word production was the only predictor that significantly contributed to the model. Word comprehension and gesture were not significant contributors. However, the sensitivity was still not good (50%), and the PPV was only 17.6%, leading the researchers to conclude that screening children at 18 months on a parental report of vocabulary was not an effective procedure. This is in accord with Henrichs and colleagues' study, which reported low sensitivity using vocabulary measures at 18 months, although the sensitivity value was even lower in Henrichs and colleagues' study (30%) despite the fact that children were followed only up to 30 months. A possible explanation for the differences in results between the two studies is the different measures used at outcome. Henrichs and colleagues used a vocabulary measure at outcome which showed a slight ceiling effect as noted by the authors, while Westerlund and colleagues' outcome measure focused on assessing children's receptive and expressive language skills. Children might have shown improvement in their vocabulary but problems in receptive and expressive language were still evident, thus leading to more children identified as delayed in Westerlund and colleagues' study. Rescorla (2000, 2002) showed that performance of late talkers at outcome varies depending on the measure used.

In the US, Thal (2005a, 2005b) as cited in Ellis and Thal (2009) also examined the predictive value of expressive vocabulary, receptive vocabulary and gestures, although late talkers were identified at a younger age (16 months) compared to the previous studies. The MacArthur Communicative Development Inventory (CDI; Fenson et al., 1993) was administered to 1,100 children at 16 months of age and children were then followed up at 28 months. Children were identified as either late producers or late comprehenders at 16 months of age based on parental reports of vocabulary. Late producers were those delayed only in vocabulary production and late comprehenders were children who were delayed in both comprehension and production. Continued vocabulary or grammar delay at 28 months of age was predicted by a combination of factors including family history of language impairment, lower gesture use on the CDI, and the identification of the child as late comprehender. Sensitivity, specificity and negative predictive values were high (0.80,

0.93, and 0.99 respectively). On the other hand, positive predictive value was low (0.16). The likelihood ratio (LR) was also calculated in this study. It was reported that children who met the above criteria at 16 months were 11.3 times more likely to show expressive language delay at 28 months. When a group (N = 577) of these children was followed up at 6 years, 8.6% of late comprehenders were diagnosed with SLI by a speech language pathologist in comparison to 3.7% percent of late producers. Both figures are strikingly small indicating that nearly 90% of children who were identified as late talkers at 16 months (late producers or late comprehenders) were classified as typical children at 6 years of age. Furthermore, more children diagnosed with SLI at age 6 had typical language histories at intake. Hence, children's language status at 16 months of age was not a reliable index for outcome in school age children. In line with the results of the previous two studies, assessing language skills in children younger than 2 years old appears to have limited predictive value.

More recently, in a retrospective study, Poll and Miller (2013) investigated whether children who showed poor language skills at 8 years of age were identified as late talkers at 2 years of age. At age 8 measures of language and cognitive skills were drawn from 1015 children. Participants only included children without cognitive disabilities based on scores on a direct measure of intelligence. Children were identified with weak language ability if they scored 1 SD or more below the mean on at least 2 of the 4 administered language tests, or 2 SD or more on 1 language test. The language assessments included naming pictures, recalling sentences, defining words and a narrating task. Seventy two children out of 1015 children met that criteria and were classified in the weak language group. A control group of 241 children of the same age were randomly selected. Late talking status at 2 years of age was identified by a vocabulary score at or below the 10th percentile on the CDI or by the absence of word combinations based on parental report. To determine whether late talking predicted poor language outcome 6 years later, logistic regression was performed. Results showed that only the absence of word combinations contributed significantly to the model when children's cognitive abilities, maternal education and race were entered as covariates. Children not yet combining words were reported to be 2.8 times more likely to be in the weak language group at 8 years of age. The authors argued that vocabulary scores may be more associated with SES and nonverbal cognitive abilities compared to the ability of combining words. Thus, relations

between vocabulary and language outcome becomes non-significant when those factors are accounted for.

1.1.2.3 Summary of large scale studies

Large scale late talker studies have proposed and investigated a number of predictors of language outcome. This section summarises findings to date on these predictors to determine their relative strength and the weight of evidence for this. However, it is important to take into account that studies have differed in several aspects such as the inclusion criteria for participants, measures used at intake and outcome, age of children included and the time interval between when the children were first assessed and follow-up. Another important factor that could affect results is whether children received therapy between baseline and outcome. The only study that reported intervention information was Westerlund and colleagues' study in which it was stated that participants did not receive therapy. Accordingly, these findings must be interpreted with caution.

Collectively studies have suggested that measures of expressive vocabulary alone are inadequate in predicting the child's longer term outcome since only a small amount of variance in outcome in preschool age is accounted for by performance on these assessments and estimates of accuracy of prediction are generally low. Predictiveness of outcome using these measures appears to be worse over longer periods between initial and follow-up assessment (Thal, 2005a, 2005b). This was also found when a retrospective design was used (Poll & Miller, 2012). Furthermore, they are less reliable in children younger than 24 months as higher rates of recovery (71%) at 30 months of age were reported in the 18 month old participants in Henrichs et al.'s study in comparison to the recovery rate (56%) of 2 year old participants in Dale et al.'s study. In addition, assessing children's ability to combine two word sentences did not seem to contribute to predicting outcome. As argued by Dale and colleagues a grammar measure may be insensitive at 2 years as most late talkers are not yet combining words.

With regard to receptive language, it appears that this has low predictive value in children younger than 2 years (Henrichs et al., 2011; Thal, 2005 a, 2005 b; Westerlund et al., 2006). It must be noted however that in all reviewed studies that examined the role of receptive language a parental report was used. Some researchers questioned the reliability of receptive language scores based on parental reports and pointed out that receptive

language is generally more difficult to judge than expressive (Eriksson, Westerlund, & Berglund, 2002; Feldman et al., 2000).

Gesture was only investigated in two studies which reported conflicting findings (Thal, 2005a, 2005b; Westerlund et al., 2006). While Thal concluded that communicative gestures were among the best predictors of continued delay, Westerlund and colleagues reported that gesture use did not predict outcome. These studies differed in the criteria used to define language delay and ages of participants at intake and follow-up (16 to 28 months in Thal, and 18 to 3 years in Westerlund et al.). In addition, Westerlund and colleagues used a short version of the CDI which contained only 12 gestures, while Thal used the full form of the CDI which contains a larger number of gestures. A previous study (Eriksson, Westerlund, & Berglund, 2002) reported low variance of scores on the gestures scale on the short version of the CDI leading to poor differentiation between children performing at the bottom and top of this scale. This may account for the finding that gestures were not a significant predictor of outcome in Westerlund and colleagues' study.

1.1.3 Small scale studies

1.1.3.1 Introduction

The demographic characteristics of small scale studies may be less representative of the broader population in comparison to large scale studies, with samples of some of these studies biased to middle high SES or low SES. Children were in some studies referred by their parents in response to advertisements specifying the age range and language level of the child while in others they were referred by their parents or teachers because of a concern regarding their expressive language development. In general, small scale studies used direct methods of assessment at intake and more stringent inclusion criteria for participants than large scale studies. For example, only children who showed normal intellectual ability based on results of direct assessment were included.

1.1.3.2 Findings of small scale studies

In two studies of a small number of late talkers, Thal and colleagues (1991, 1992) examined predictors of persistent language difficulties in 10 late talkers aged 18-29 months who were compared with age matched and language matched typically developing

children. Children were identified as late talkers if their scores were in the lowest 10% on a parental report of expressive vocabulary. In addition, information was collected on children's comprehension based on a parental report; a two-way forced choice picture identification task; mean length of utterance (MLU) in a spontaneous one-hour language sample; and gesture use as measured on an experimental imitation task. Participants were followed up after one year. Based on vocabulary production scores reported by parents, the authors reported that 6 children appeared to have caught up while 4 children remained delayed. Results showed that children's language production scores did not differentiate between the groups, whereas language comprehension and gesture use scores were significantly lower for children who remained delayed at follow up (Thal, Tobias, & Morrison, 1991). However, caution must be exercised in interpreting these findings since the sample size was small.

In the (1992) study, Thal and Tobias transcribed and analyzed gestures used by the children in the 1991 study during the one hour videotaped interaction sample. They found that children who caught up at the one year follow-up had produced significantly more communicative gestures at intake in comparison to those who were truly delayed. These results led the authors to suggest that late talkers who recover use communicative gestures to compensate for their limited verbal production at time 1.

Recently, Bishop and colleagues (2012) reported 4-year-old outcome for 24 late talkers identified at 18 months with a score of more than 1 SD below the mean on the Oxford University Communicative Development Inventory (OCDI; Hamilton, Plunkett, & Schafer, 2000). A control group of 58 typically developing children (TD) who scored between the 30^{th} and 75^{th} centile on the OCDI were also recruited. Parents completed the receptive and expressive vocabulary components of the OCDI at 18 months and children were invited for a direct assessment 4 to 6 weeks later. During the direct assessment the receptive and expressive parts of the Mullen Scales of Early Learning (MSEL; Mullen, 1995) and the Vinelands Adaptive Behavioral Scales (VABS; Sparrow, Cicchetti, & Balla, 2005) were administered. Children were followed up at 4 years of age and the following abilities were tested in a direct assessment: comprehension, expressive vocabulary, phonology, grammar, narrative ability and intellectual ability. In addition, the child's communicative ability was assessed using a parental report and parents were invited to complete a nonword repetition task. A diagnosis of SLI at follow-up was defined as impaired performance on two or more of the nine administered language measures with

average nonverbal ability. More than two thirds of late talkers (62.5%) were in the typical language group at 4 years of age. A step-wise discriminant function analysis that included children's receptive language, expressive language, communicative ability, parental nonword repetition and family history of language-literacy problems showed that children's receptive language obtained from direct assessment and the parental nonword repetition score were significant predictors of outcome with high specificity (0.98) but low sensitivity (0.46). The authors concluded that poor comprehension should be viewed as a 'red flag' suggestive of persisting difficulties especially when measured by a direct assessment. However, they also suggested that it might be too early to screen children who are younger than 2 years of age for language problems.

Rescorla and colleagues conducted the longest late talker follow-up studies to date (Rescorla, 2002, 2005, 2009; Rescorla, Dahlsgaard, & Roberts, 2000; Rescorla, Roberts, & Dahlsgaard, 1997). They examined the outcome for late talkers who were identified at 24-31 months and then followed up at different points during preschool and school up to 17 years of age. The original sample included 40 late talkers and 39 typically developing children matched on age, socioeconomic status, and nonverbal ability (Rescorla et al., 1997). Late talker inclusion criteria at intake included normal receptive language and a score at least 6 months below chronological age on the Reynell Expressive Language Scale (Reynell, 1977). All late talkers met Rescorla's (1989) cut-off for language delay (fewer than 50 words or no word combinations).

Follow-up data at age 3 and 4 years on the Index of Productive Syntax (IPSyn; Scarborough, 1990) and mean length of utterance (MLU) showed that late talkers made larger gains on these measures in comparison to matched controls (Rescorla, Dahlsgaard, & Roberts, 2000). However, 66% of the late talkers scored below the 10^{th} percentile on IPSyn at age 3 and 71% did so at age 4. MLU scores showed that 59% of late talkers were still delayed at age 3 but only 29% at age 4. The authors concluded by suggesting that the rate of recovery of late talkers in terms of MLU is about 50% per year from age 2 to age 5. They further noted that children who are still delayed at 5 years are likely to have persistent language problems.

By 6-9 years, late talkers generally scored in the average range on most language measures (Rescorla, 2002). Nevertheless, the percentage of children scoring below the 10^{th} percentile was usually higher in the late talker group compared to the control group. In the

late talker group the percentage ranged from 0% to 68% and in the control group it ranged from 0% to 41%. On 5 out of the 29 measures all late talkers performed in the normal range and only on 2 measures the percentage of late talkers performing below the 10th exceeded 40%, while the control group performed in the normal range on 20 out of the 29 measures.

At 13 years of age, Rescorla (2005) reported that most late talkers scored in the normal range on various language measures although significantly below their peers. In addition, the Language Development Survey (LDS; Rescorla, 1989) score at intake was a significant predictor of outcome measures at age 13 including vocabulary, verbal memory and reading comprehension, with the variance explained ranging from 14% to 20%. On the other hand, receptive language score at two years was not a significant predictor of outcome despite the fact that there were significant differences between late talkers and the control group on the receptive language score. This may be explained by the fact that only late talkers with normal receptive language were included.

Similarly, Rescorla's (2009) follow-up study of children at age 17 showed that most late talkers scored in the average range on language and memory measures although more poorly than the control group. Stepwise multiple regression indicated that age 2 LDS vocabulary score explained 17% of the variance of age 17 vocabulary and grammar factors. Reynell expressive and receptive scores added only 3% and 1% respectively. In conclusion, the author suggested that the findings of the developmental continuity of language abilities supports the dimensional account of language skills according to which typically developing children, late talkers, and children with receptive and expressive language delays differ quantitatively on a hypothetical language spectrum. Late talkers with expressive delays only may perform below peers on language measures at follow-up because they have weaker endowment of the skills that sub-serve language, late talkers who have receptive and expressive delays have a more compromised endowment and may be more impaired at follow up. On similar lines, Desmarais Sylvestre, Meyer, Bairati, and Rouleau (2008) in a review of late talker studies drew attention to the heterogeneity of language profiles of late talkers and highlighted the need to collect information on the progression of late talkers grouped according to several early prelinguistic or linguistic communication skills to help guide clinicians in deciding when immediate intervention is indicated. The authors also emphasized the important role that early receptive language

may play in development and the need for future research to investigate whether 2-year-olds with comprehension deficits are at greater risk of persistent difficulties.

Recently, two groups of researchers have widened the search for predictive factors of persistent language and communication difficulties in young children with language delays (Chiat & Roy, 2008, 2013; Everitt, Hannaford & Conti-Ramsden, 2013). Participants in both studies were older than the children in the previous large and small scale studies. These samples differed in another respect: in both of these studies children with language delay were recruited because someone had concerns about their language development, unlike the late talking children in other studies which formed the group scoring below the 10th percentile for age on a vocabulary or expressive language measure. In other words, the groups of children with language delay in these two studies might have more severe problems in comparison to typical late talker samples. These differences need to be taken into account in comparing findings.

Chiat and Roy (2008) investigated whether measures of very early processing skills that include social responsiveness, joint attention, symbolic comprehension and word/nonword repetition predict the likelihood and nature of persistent language and communication difficulties in children referred at 2;6-3;6 years. Their sample of 187 participants, were children referred to clinical services with concerns about language. At intake, children's language was assessed using the UK short version of the MacArthur Communication Development Inventory (MCDI-UKSF; Roy, Kersley, & Law, 2005) and the Pre-school Language Scale- 3 (UK), Auditory and Expressive (PLS; Boucher & Lewis, 1997). Children's sociocognitive and phonological skills were assessed using the novel Early Sociocognitive Battery (ESB, Chiat & Roy, 2006b) and Preschool Repetition Test (PSRep; Seeff-Gabriel, Chiat, & Roy, 2008) assessments respectively. After 18 months, 163 children were followed up and reassessed on the PLS-3 UK. In addition, measures of morphosyntax and social communication were administered. Regression analysis was used to identify predictors of outcome. Results showed that receptive and expressive scores on the PLS-3 UK, sociocognitive composite and word/nonword repetition score on the PSRep all predicted expressive language at outcome. Furthermore, receptive language was the most powerful predictor of persistent language difficulties, accounting for 29-39% of the variance in language measures. But importantly, the measure of sociocognition was the best predictor of social communication outcome and the measure of phonology was the best predictor of morphosyntactic outcome. On the basis of these findings the authors

suggested that the assessment of these early processing skills contributes not only to the prediction of longer term difficulties but also to the nature of these difficulties, with important implications for early intervention. Likewise, Jansen and colleagues (2013) have emphasized the importance of using a dimensional approach for language assessment and underlined the importance of including assessments of joint attention and symbol understanding for young children with language difficulties. In this study, cluster analysis was used to group children with language difficulties ranging in age from 24 to 46 months into homogenous subgroups. Results suggested that in addition to cognitive functioning and the presence of autism spectrum disorder related characteristics, the mastery of the prelinguistic skills of joint attention and symbolic comprehension differentiated between the subgroups of children.

In a further follow-up 7 years later, Chiat and Roy (2013) reassessed 108 children from their original study, now aged 9-11 years, on a number of language and social communication measures. Logistic regression analysis revealed that the Auditory PLS was a predictor of all outcome measures which included language, social communication and morphosyntax, with adequate levels of specificity (0.77-0.84) but low sensitivity (0.5-0.57). The ESB was the strongest predictor of social communication problems as measured on parental report using the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005). Results showed high specificity (0.89), high positive and negative predictive values (0.7 and 0.83 respectively), but low sensitivity (0.57), leading the authors to suggest the need to investigate other predictors of social communication deficits. Interestingly, these results demonstrated that the ESB did almost as well as the Auditory PLS in predicting language outcome. On the other hand, word/nonword repetition was predictive of nonword repetition but no longer a specific predictor of morphosyntactic outcome. The authors proposed that these findings might be the result of the baseline sample including many children with severe phonological production problems which affected their performance on the PSRep but did not have longer term implications for language development.

Similarly, a recent study by Everitt, Hannaford and Conti-Ramsden (2013) searched for markers of persistent expressive language delay (PELD) in children aged 3-4 years whose teachers expressed concerns about their expressive language. This study focused on investigating linguistic skills known to relate to persistent language difficulties. Their study included a group of 47 children with specific expressive language delay and 47

typically developing children matched for age and sex. The sample tended to be from lower SES with a fairly high proportion of parents with no educational qualifications. At baseline, children's language was assessed using the PLS-3 (UK) and test markers including recalling sentences, word/nonword repetition, digit recall and a tense task. Children were followed up 12 months later. Based on their scores on the PLS-3 (UK) Expressive Communication (EC) subscale, children were classified into typical and persistent language difficulties groups. Results showed that scores on the receptive and expressive subscales of the PLS-3 UK and recalling sentences were predictors for risk of PELD. In addition, taking performance at the 16th centile on recalling sentences as a marker of future caseness threshold yielded high sensitivity (.95) and specificity (.81), reasonable positive likelihood ratio (LR+ = 5.11) but low negative likelihood ratio (LR- = 0.07). However, unlike Chiat and Roy's (2008) study, word/nonword repetition only differentiated between groups at baseline and did not predict outcome. The authors argued that the differences in the age of the children and recruitment criteria between the two studies might explain these inconsistent findings. Furthermore, differences between samples in participants' SES may have been a significant factor. More importantly, the two studies differed in how they examined the predictive value of word/ nonword. While Chiat and Roy considered also the nature of longer term difficulties and examined word/ nonword repetition as a predictor of morphosyntactic outcome, Everitt and colleagues only measured general language outcome.

1.1.3.3 Summary of small scale studies

Like large scale studies, different factors which are proposed to place late talkers at risk have been investigated in small scale studies. This section will summarize findings of these studies to determine the most reliable predictors of outcome. It should be kept in mind however that the findings must be interpreted with caution due to the substantial differences in studies designs and methodologies. Furthermore, a small number of participants were included in most studies – apart from Chiat and Roy's study - the total sample size ranged from approximately 20 participants to 90 participants (late talkers ranging from 10 to 47).

Three research groups investigated predictiveness of expressive vocabulary. Thal and colleagues (1991) and Bishop and colleagues (2012) reported that children's expressive vocabulary score did not differentiate between truly delayed children and late bloomers. Late talking children in the two studies were 18 months old at baseline assessment.

Accordingly, Bishop and colleagues suggested that 18 to 20 months may be too early for identifying children with risk for language difficulties. This is in line with the conclusion drawn by researchers in large scale studies (section 1.1.2.3) who questioned the effectiveness of screening children younger than 24 months old. The predictive role of expressive vocabulary was also examined by Rescorla and colleagues. They reported that parental report of expressive vocabulary at 2 years of age was a significant predictor of language outcome at 13 and 17 years of age, accounting for around 14%-20% of the variance of measures of vocabulary and grammar (Rescorla et al., 2005, 2009). This finding though interesting, is not very important in terms of clinical application as most late talkers in Rescorla's studies performed within the normal range on most standardized assessments and were not identified as language impaired at follow-up. Thus, it appears that vocabulary measures alone have limited predictive value for group membership at outcome.

Turning to receptive language, unlike the findings of large scale studies, overall results of small scale studies have shown that receptive language is among the best predictors of outcome (Bishop et al., 2012; Chiat & Roy, 2008, 2013; Everitt et al., 2013; Thal et al., 1991). This is likely due to mostly using direct measures in assessing receptive language and not parental reports as the case in the large scale studies. In addition, children in the Chiat and Roy's study and Everitt and colleagues' study were older and might have shown more severe problems. Whilst Rescorla's (2005, 2009) studies did not support this conclusion, it's important to reiterate that her studies recruited children with normal receptive language.

Expressive language was also suggested to be a significant predictor in Chiat and Roy's (2008) and Everitt and colleagues' (2013) studies but it was not a significant predictor in the younger group (18 months-20 months) in Bishop and colleagues' (2012) study. Again, as mentioned above, the samples in these studies differed not only in the age range but also in the recruitment procedure.

Another finding in the reported studies is the heterogeneity of late talkers and the necessity to assess individual differences in language-related abilities such as gestures (Thal et al., 1991, 1992) and the sociocognitive skills measured in Chiat and Roy's study. It should be noted that the gesture tasks in Thal and colleagues' (1991) study largely involved imitation

skills rather than communicative gestures as children were asked to imitate the different gestures modeled by the examiner.

Other early processing skills have been put forward as candidates to assist prediction of continued language impairment. Although word/nonword repetition appears to be a good marker in discriminating between young children with and without language difficulties, sentence repetition seems to be a better predictor of longer term difficulties (Everitt et al., 2013). However, investigation of these predictors is limited to date for the very young age range under consideration.

The next section will summarize findings of a study that used a parental report of children's language use - the Language Use Inventory (LUI; O'Neill, 2009) - to predict language outcome. This study as mentioned previously used a very different methodology in comparison to large and small scale studies and for that reason it is reviewed in a separate section.

1.1.4 Findings of the LUI study

Pesco and O' Neill (2012) examined the ability of a new parental report of children's use of language -the LUI- to predict language outcome in 348 children between 18 and 47 months. The development of the LUI was motivated by the knowledge that some children's communication difficulties are mainly manifested at the level of pragmatics (Bishop & Norbury, 2002); that is, for some children, impairments are shown in their abilities to use language effectively and appropriately in interacting with people in everyday settings (Bishop & Norbury, 2002). The LUI measures a number of communicative functions. It includes items on child's ability to direct attention, ask and comment about things and themselves, talk about people's behaviors, use mental state terms and build longer sentences. To assess the predictive validity of the LUI the investigators employed a methodology that differs from small and large scale studies in many respects. The large number of participants was recruited from a database of children who participated in the norming study of the LUI, but selective sampling was used to ensure that an adequate number of low scoring participants were included in the sample. Most of the children came from homes with middle to high income. Inclusion criteria for participants were more stringent than in other large late talker studies with only children who were born full term, not exposed to a language other than English more than 20% of

waking hours, and not diagnosed with hearing loss or speech language delay included. Children were followed up at 5-6 years and reassessed using three standardized measures of language and a parent report of developmental history. A delay at outcome was identified as a score 1.5 SD below the mean on any of the three administered language tests or a history of a diagnosed language delay, language impairment, or autism as reported by parents. For the youngest age groups (18-23 months), results showed that the majority of children who scored below the cutoff at intake were in the normal group at outcome. In line with results of previous sections this finding indicates that assessing children who are younger than 2 years old have limited predictive value. With regard to children aged 24-47 months, high sensitivity (0.81) and specificity (0.93) were obtained at the 5th percentile cutoff. A child scoring below the 5th percentile had 27 times greater risk of showing later language difficulties at age 5-6 years. This finding is interesting given that the children at follow-up in Pesco and O'Neill's study (mean age 5;8 years) were older than children at follow-up in most large scale late talkers' studies reviewed in section 1.1.2. Thus, the LUI successfully predicted outcome up to school age. However, it must be noted that a wider range of assessments were used at outcome in Pesco and O'Neill's study. Consequently, more children could be classified as impaired at outcome in comparison to other studies. Furthermore, the nature of language or communication problems identified at outcome was not reported. In addition, some of the children at baseline were older than children in late talker studies and interval between time 1 and follow-up varied among participants from 14.54 to 54.76 months. Moreover, the sample included very few children from a disadvantaged background thus children falling below the cutoff might show quite severe problems relative to a more diverse sample. These differences between the studies may partly account for better accuracy estimates of the LUI. Accordingly, replicating these findings in further research with a more diverse sample could provide more definitive evidence on the predictive value of the LUI.

To my knowledge, no other study has investigated whether assessing young children's use of language for different pragmatic functions would predict later language outcome, although previous studies with small numbers of participants have suggested that late talking toddlers initiated joint attention less often than typical age matched peers even through nonverbal means (Paul & Shiffer, 1991), asked fewer questions, and produced fewer declarative statements (Rescorla, Bascome, Lampard, & Feeny, 2001). Given the

high sensitivity and specificity values of the LUI it appears to be an effective means to assess early pragmatic functions in toddlers.

1.1.5 Conclusion

This chapter has reviewed a number of longitudinal studies of young children with early language delay to identify language and communication measures that have been found to be informative of outcome. Despite the large variations in design and methodology a number of key findings can be drawn.

There is replicated evidence that assessing late talkers who are younger than 2 years of age has limited predictive value for later outcome. This finding is almost consistent in both large and small scale studies. Accordingly, in the current study, the minimum age of participants was 2 years. Examining results of the studies further highlighted the heterogeneity of late talkers and the need to assess individual differences in several skills as relying only on measures of expressive vocabulary was found to be inadequate. Assessing sociocognitive skills such as gesture use has been suggested to be useful for predicting of continued language delay (Thal et al., 1992) but this was based on a very small sample of late talkers followed up after 1 year. The sociocognitive skills assessed in the ESB have been shown to contribute distinctively to specific problems of social communication 7 years later. This finding suggests that this measure is not only informative about persistent problems but also has the potential to predict the nature of problems which is important in providing targeted intervention.

Children's receptive language was found to be strongly related to outcome. Children with receptive language delays were more likely to have persistent language problems at outcome especially when receptive language was assessed using a direct method of assessment and not a parental report.

The LUI emerged as an unusually strong indicator of language outcome in Canadian children at 5-6 years of age with excellent sensitivity and specificity values. This finding is of interest particularly as, according to the guidelines of major organizations such as the American Speech and Hearing Association, the assessment of the different functions of communication is an important area to consider when evaluating children with early

language delay (ASHA, 2008; Crais, 2011). Given the excellent sensitivity and specificity values of the LUI it appears to be an effective measure of communicative functions.

In line with the recent interest in using processing markers to identify and predict SLI, two studies have recommended the assessment of early processing skills such as word/nonword repetition and sentence repetition as predictive of continued language impairment in young children with language delays. Sentence repetition was a better predictor of outcome in children between 3-4 years old. Given that there is a sentence repetition test that is developed and normed for Saudi children (Wallan, Chiat, & Roy, 2011) the inclusion of such a measure would also provide an external validation for newly developed or adapted measures in this study.

To conclude, this project aimed to develop a battery of measures to assess children aged 2;0-3;5 years. For the development of the battery it was important to include measures that help with not only identifying young Saudi children with language and /or communication problems but also understanding the nature of their problems. Based on the above findings, this will include the ESB, a direct measure of receptive language, an adaptation of the LUI and a sentence repetition test. Not only do these measures have strong predictive value but also the combination of the measures is in line with the recommendation to gather information about the child's use of language in different contexts, and to include multiple types of tests such as parental reports and direct assessments when assessing young children with language delays in order to obtain a more holistic picture of the child's language skills (Crais, 2011).

In line with this view, two further measures were included in this battery: a motor imitation measure and sociocognitive questionnaire. Motor imitation has commonly been proposed to be an important prelinguistic skill relating to language and communication development in typically developing children and children with autism though less attention has been paid to including a test of motor imitation as a predictor of outcome in late talker studies, with only one study (Thal et al., 1991) suggesting differences in late talkers' outcome based on their imitation of gestures. Including a measure of motor imitation might provide a better understanding of the child's strengths and weaknesses and improve the predictive value of the assessment battery. Furthermore, using a parental report of sociocognitive skills may contribute to the assessment by providing information especially on skills that are difficult to assess in a clinical setting.

CHAPTER 2

SOCIOCOGNITIVE SKILLS AND LANGUAGE

2.1 Introduction

Social cognition "refers to the ability to understand other people" from the ability to use social cues such as eye gaze, facial affect, and voice tone to the ability to use symbolic communication (Striano & Reid, 2006, p.471). A main focus in this project was on assessing early sociocognitive skills that children use to understand the communicative intents of others. These basic skills usually emerge at about 9-12 months. During this period infants engage in triadic interactions with others and use others gaze, point, facial and vocal affect to infer the speakers' referential intent. These nonverbal sociocognitive skills have been described as fundamental abilities to understand the communication intention which form the foundation for language development (Mundy & Gomes, 1997).

The rationale for assessing these early sociocognitive skills in the current study is grounded in the sociocognitive hypothesis which was supported by the findings of Chiat and Roy (2008, 2013). According to this hypothesis, constraints on sociocognitive skills will limit children's ability to use pragmatic cues to identify speakers' meaning intentions and hence to discover the meanings behind their words, which will affect the acquisition of language and its use (Chiat, 2001; Chiat & Roy, 2008, 2013). Accordingly, children with sociocognitive constraints are expected to show specific difficulty in language comprehension and social communication.

Early sociocognitive skills have been the focus of much research on language development in typically developing children and children with autism. However, less attention has been paid to the assessment of these skills in late talkers or children with early language delay despite the evidence of deficits in social cognition in older children with SLI and overlap between the problems observed in SLI and autism (Leyfer, Tager-Flusberg, Dowd, Tomblin, & Folstein, 2008). Only recently have researchers highlighted the importance of assessing these skills in young children with language problems (Chiat & Roy, 2008; Desmarais, Sylvestre, Meyer, Bairati & Roulea, 2008; Jansen et al., 2013). Acknowledging the importance of systematically assessing sociocognitive skills Chiat and Roy (2006a, 2008) developed the Early Sociocognitive Battery (ESB). The ESB assesses

the skills of social responsiveness, joint attention and symbolic comprehension. It focuses on assessing these sociocognitive skills using tasks that measure responses to input rather than elicitation tasks. It is a very quick measure that has been shown, as mentioned in the previous chapter, to be a strong predictor of social communication problems 7 years later with high specificity (0.89) and high positive and negative predictive values (0.7 and 0.83 respectively) though low sensitivity (0.57). Other similar measures that assess social cognitive skills either take longer to administer and score such as the Early Social and Communication Scales (ESCS; Mundy et al., 2003) which focuses mainly on assessing different aspects of joint attention and behavioral regulation; target children younger than 24 months of age such as the Communication and Symbolic Behavioral Scales-Developmental Profile (CSBS-DP; Wetherby & Prizant, 2002); or assess fewer aspects of social cognition. Furthermore, the predictive validity of most of these measures has been assessed only up to preschool or young school-aged children. For example, the CSBS predictive validity was assessed based on outcome when children were 3 or 4 years of age. The ESB is not intended to assess all aspects of sociocognition. Nevertheless, it is quite comprehensive and assesses skills that are considered necessary for social communication development.

This chapter will briefly explore the literature relating to these three skills of sociocognition focusing on their relations with language. Another sociocognitive skill that has been highlighted in studies of typically developing children and the literature on autism and was developed as part of this project is motor imitation. The literature on motor imitation will be the focus of the next chapter.

2.2 Social Responsiveness and Language

The first task in the ESB focuses on the child's attention to the experimenter's expression of emotion. Affective communication is an integral part of infants' early communication with other people. Researchers have shown that, from the early months of life, infants appear to attend to faces and show differential responsiveness for different facial and vocal expressions (Montague & Walker-Andrews, 2001). The participation of infants in face to face dyadic interactions has been termed "primary intersubjectivity". These emotional exchanges have been suggested to represent the foundation for later social competencies (Clifford & Dissenayake, 2009) and emotional development (Stern, 1985). A key

transition in development happens when infants' affective exchanges move beyond the context of dyadic interaction and become referential towards objects or events (Bates, 1979). This triadic interaction has been termed "secondary intersubjectivity" (Trevarthen & Hubley, 1978). During this period infants check adults' emotional perspective to guide their behavior and disambiguate a novel situation, an ability termed 'social referencing' (Walden & Ogan, 1988). In addition, by 18 months of age, infants can regulate their own actions on the basis of their memory of their observation of emotional reaction directed at an experimenter and not themselves (Repacholi & Meltzoff, 2007). After observing an adult expressing anger towards an experimenter playing with an object, infants were hesitant in playing with the object in this study.

Furthermore, in a subsequent study, Repacholi, Meltzoff, and Olsen (2008) showed that infants are able to integrate several sociocognitive cues. In this study not only did infants appear to regulate their behaviors based on their memory of the previously angry adult, but they also appeared to check the head/eye direction or eye status of the previously angry adult and integrate this information to regulate their own behavior and decide whether or not to imitate an adult in playing with an object. Thus, they appeared to appreciate that others' gaze direction affects their emotional reaction. More recently, Repacholi, Meltzoff, Rowe and Toub (2014) showed that even 15-month-old infants were able to integrate emotional and visual-perceptual cues to regulate their imitative behavior.

Infants' sensitivity to emotional expressions also plays a role in guiding word learning. Two early studies (Tomasello & Barton, 1994; Tomasello, Strosberg, & Akhtar, 1996) demonstrated that 1;6 and 2;0 year old children were able to learn new words in a searching game by using the examiner's facial and vocal expressions to determine the examiner's intended referent.

More recently, Berman, Chambers and Graham (2010) and Berman, Graham, Callaway and Chambers (2013) have shown that preschool children understand the meaning behind different vocal affects and learn new words based on information carried by using vocal cues. Using eye tracking the researchers showed that children were more likely to look at a broken object when a spoken statement was produced using a negative affect.

In contrast there is replicated evidence that children with autism are less attentive to others' emotions (Charman et al., 1997; Sigman et al., 1992). Their deficits in using these

important pragmatic cues and the difficulties they have in processing emotional information might impede their pragmatic development (Dawson et al., 2004).

While most of the early studies that have examined attention to distress in children with autism were cross sectional, and involved children who had already been diagnosed, recent longitudinal studies have suggested that low social responsiveness to distress may indicate a higher risk for autism (Hutman et al., 2010). In this study, the researchers recruited infant siblings of children with autism and low risk infants with no family history of autism. Infants' responsiveness to distress was assessed at 12, 18, 24 and 36 months. Infants who were later diagnosed with autism showed less attention and affective responsiveness to distress than comparison groups. Furthermore, the researchers showed that response to distress was related to language skills.

Although the literature focuses on deficits in emotional relatedness in autism, there is evidence that school-age children with SLI may show deficits in emotional understanding (Fujiki, Spackman, Brinton & Illig, 2008; Spackman, Fujiki, & Brinton, 2006; Taylor, Maybery, Grayndler & Whitehouse, 2015). However, most of these studies approached this issue from a different angle as the assumption was that language impairment affected emotional development and not the other way around.

Given that some difficulties have been reported on different tasks of emotional understanding in school-age children with SLI it is reasonable to argue that some of these children may show impairment in basic level of emotional engagement at an earlier age. In support of this suggestion a recent longitudinal study (Hutman, Rozga, DeLaurentis, Sigman, & Dapretto, 2012) showed that typically developing infants' attention to distress at 12 months was related to children's receptive and expressive language at 36 months, explaining approximately 4% and 5% of the variance respectively. In addition, children who showed congruent affective responses at 12 months had marginally higher receptive language scores at 36 months. The authors concluded that "more responsiveness to others' distress corresponds with greater aptitude for language learning from social interaction" (p.10). Although this study showed that attention to positive emotions was not related to their language, it is unclear how the positive emotions were expressed during the interaction. It is possible that infants are more attentive to sudden or unexpected changes in emotions during an interaction and not necessarily negative emotions.

2.2.1 Social responsiveness and language conclusion

Collectively, these findings are compatible with the argument that the use of emotional cues is crucial to discovering meaning intentions behind utterances and hence meaning of speaker's words. However, it is not sufficient and further sociocognitive skills are needed. Notably, the ability to follow or check other's gaze and point to determine their focus of attention have been argued to be crucial for early language development. A brief overview of the studies that examined relations between joint attention and language is presented in the next section.

2.3 Joint Attention and Language

2.3.1 Introduction

The second task in the ESB assessment is joint attention. Joint attention refers to the ability to coordinate attention between two individuals in relation to objects or events for social purposes (Mundy, Sigman, Ungerer, & Sherman, 1986). Researchers have investigated different behavioral manifestations of initiating or following joint attention such as gaze following, point following, pointing gestures, and gaze alternation. Most scholars agree that joint attention involves not only a child and an adult focusing on the same thing such as synchronized looks to a noisy object, but also the notion that both are aware that they are sharing the focus of attention (see Carpenter & Liebal, 2011). Acts of joint attention have been described as an entry point for understanding people's minds (Brooks & Meltzoff, 2014) and considered important social cognitive milestones that are interrelated with other social cognitive components (Carpenter et al., 1998) and central to early word learning (Tomasello, 1995, 2001).

Relations between joint attention and early language have been the focus of much research. The following sections will provide a brief overview of studies that investigated how joint attention contributes to language learning and the predictive role it plays in relation to language development in typically developing children and children with autism spectrum disorders (ASD). Though different types of joint attention have been considered important elements of the shared experience that is necessary for language acquisition (Tomasello, 1995) and correlational relations have been found between the different types of joint attention and language (Luyster, Kadlec, Carter, & Tager-Flusberg, 2008), the focus of this review will be on studies of "*responding to joint attention*"(RJA).

RJA refers to the ability to follow the gaze and or the pointing gesture of others (Morales et al., 2000). This ability provides infants with important cues to the intended referents of others' language and hence provides a base for the word mapping process. Accordingly, based on the sociocognitive hypothesis, impairments in responding to joint attention will impact on the word mapping process and the ability to understand the meaning of words. The studies reviewed in the following sections examined relations between RJA and language using a variety of approaches. In one approach, researchers manipulated the contextual attentional cues for novel word learning to explore how these would affect children's ability to learn a novel word. Studies using this approach are reviewed in section 2.3.2. Alternatively, researchers have examined the association between individual differences in RJA and later language in typically developing children or children with autism spectrum disorders (ASD). These studies are reviewed in sections 2.3.3 and 2.3.4.

2.3.2 Experimental studies of responding to joint attention and word learning

Much research has focused on how children learn what words refer to through acts of joint attention, mainly through gaze following. One of the earliest experimental studies showing that children learn words more easily in the context of joint attention was conducted by Tomasello and Farrar (1986). In this study an adult tried to teach infants (mean age = 17 months) four new words using two different strategies to establish joint attention. They found that children were more successful in learning new words if the adult followed in the child's focus of attention and introduced the word when the child was focusing on the target object than when the adult tried to redirect the child's attention to the target object. This was tested in a comprehension task that showed that infants performed better in identifying the objects taught in the follow-in condition. However, by 18 to 19 months old, infants were shown to play an active role in word learning and were able to shift their attention to match it to what an adult was attending to while labeling an object (Baldwin, 1993). On the other hand, children with autism who are known to have deficits in joint attention (Sigman & Ruskin, 1999) show a tendency to make incorrect links between words and referents (Baron-Cohen, Baldwin, & Crowson, 1997). Baron-Cohen and colleagues suggested that children with autism (mean age = 9.2 years) in their study made mapping errors as a result of their inability to shift their attention to match the speaker's gaze direction. Instead they linked the new word to the object they themselves were looking at when the word was uttered. Findings of a recent study that used eye-tracking

supported the suggestion that word learning difficulties in children with ASD are related to atypicalities in gaze following (Akechi, Senju, Kikuchi, Tojo, Osanai, & Hasegawa, 2011). In this study, the performance of children with ASD (mean age = 9.1 years) in attending to objects presented on a computer screen in different naming conditions was compared to performance of typically developing children. In the discrepant labeling condition a schematic speaker face and two novel objects appeared on the computer screen, when the child focused on a particular object for 300 ms the speaker face looked at the opposite object and labeled it. Results showed that children with ASD attended equally to the target object and the distractor object in the discrepant labeling condition, while typically developing children looked at the target object longer than the distractor object. However, unlike Baron-Cohen and colleagues' study, no significant differences were found in the frequency of gaze following between children with ASD and typically developing children. Accordingly, Akechi and colleagues suggested that the difficulties in word learning were not related to the frequency but to the duration of gaze following to objects. It must be noted however that there are important differences between the two studies which may explain the inconsistent findings. First the studies differed in the context of teaching new words. Second, while the participants with ASD in Baron-Cohen and colleagues' study were relatively low-functioning and profoundly language impaired, those in Akechi and colleagues' study were high functioning and their verbal mental age was not significantly different from typically developing children of the same age. Accordingly, the differences in the findings of two studies may be explained by Luyster and Lord's (2009) suggestion that different groups of children with ASD vary in their ability to use joint attention in learning new words. Luyster and Lord further argued that due to the variability in joint attention among children with ASD it is considered a powerful predictor of later language.

To summarize, studies have shown that gaze following ability is an important skill that children use in word learning. Differences in gaze following were related to children's performance on comprehension tasks testing children's learning of novel words in both typically developing children and children with ASD. Differences in the frequency and/or duration of gaze following ability may be linked to variability in word learning ability.

2.3.3 Relations between responding to joint attention and language in typically developing children

A considerable body of evidence has documented relations between joint attention and language in typically developing children using different tasks of responding to joint attention. However, there are some inconsistencies in whether it predicted later receptive language (Beuker, Rommelse, Donders, & Buitelaar, 2013; Brooks & Meltzoff, 2005), expressive language (Carpenter et al., 1998; Delgado et al., 2002; Meltzoff & Brooks, 2008), or both (Morales et al., 2000; Mundy et al., 2007; Watt, Wetherby & Shumway, 2006). These inconsistencies are likely due to the substantially different ways of measuring joint attention and language skills. In addition, some studies assessed only expressive language. When a receptive language measure was administered, relations with joint attention were usually found to be significant. While the majority of these studies focused on assessing RJA in children between the first and second year, thus considerably younger than the children in the current study, a review of these studies will shed light on relations between different aspects of attention-following and language and the factors that influence these relations. The main findings of these studies and possible reasons for inconsistencies between them are highlighted and summarized in the following section, and the implications for the assessment of joint attention in the ESB will be considered, taking into account the differences in age. For a detailed description of measures and methods of analysis used in the different studies see Table 1.

In an early longitudinal study, Carpenter and colleagues (1998) examined relations between the age of emergence of several sociocognitive skills including joint attention and referential language (words produced spontaneously during the session) in 24 infants assessed monthly from 9 months up to 15 months. Results showed that the age of emergence of RJA – defined as the age that an infant passed either gaze-following or point-following tasks – correlated with later referential language. Relations with receptive language were not examined in this study.

A more recent study (Beuker, Rommelse, Donders, & Buitelaar, 2013) that used a methodology largely based on Carpenter et al.'s study, but spanning a larger age range, reported significant relations between following attention and receptive vocabulary. In this study a group of 23 infants were assessed monthly between 8 and 24 months, on a range of joint attention skills including tasks of gaze- and point-following to objects within or outside the visual field. Analysis showed significant correlations between the age of

emergence of following attention within the visual field at 10 and 11 months and receptive vocabulary at 18 months, and between following attention outside the visual field at 14 months and receptive vocabulary at 18 months. Children who acquired attention-following skill earlier had a larger receptive vocabulary. On the other hand, unlike Carpenter and colleague's study, relations with expressive vocabulary were not reported, and it is not clear whether this was because they were not-significant, or whether they failed to reach their criterion of 0.40 for reporting of significant correlations. Level of correlations between RJA and expressive language in Carpenter and colleagues study was less than .40 (r = .36, p < .05). In addition, it is important to note that the two studies differed in their measure of expressive language. While, Carpenter and colleagues measured the number of words produced spontaneously during the session, Beuker and colleagues used a parental report measure of expressive vocabulary.

The two studies reviewed above measured the presence/absence of attention-following skills. However, interval measures such as duration or frequency of attention-following, which are more sensitive to the variability in children's ability to follow attention, appear to be more informative about later language. Not only were significant correlations found with later language when these measures were used but unique predictive relations were also reported when the effects of cognitive ability or initial language were controlled for (Delgado et al., 2002; Morales et al., 2000; Mundy et al., 2007).

For example, using regressional analyses, Delgado and colleagues (2002) showed that differences in the ability to follow attention to targets outside the visual field at 15 months of age provided unique information about infants' later language ability. The ability to locate objects outside the visual field explained 28% and 8% of the variance of expressive and receptive language respectively at 24 months of age. However, when ability to locate objects within the visual field was controlled for significant predictive relations were found only with expressive language. The authors reported that there was a ceiling effect on attention-following trials to tasks within the visual field while only few infants were consistently capable of locating objects outside their visual field. In addition, this study showed that relations between responding to joint attention to targets outside the visual field and expressive language remained significant even when the variance of general cognitive ability was taken into account. The finding of stronger relations between responding to joint attention and expressive language in comparison to receptive language

may be explained by the fact that the measure of receptive language was reported to be less reliable than the expressive language measure (Delagado et al., 2002).

Predictive relations were also reported between RJA and language when the variance associated with initial language status was controlled for (Morales et al., 2000). In this study, 95 infants were followed from 6 to 24 months. Regression analysis showed that an aggregate measure of RJA from 6-18 months made a unique contribution to both receptive and expressive vocabulary at 30 months even when language status at 24 months had been entered. However, it is important to note that, unlike the previous studies, infants in Morales et al.'s study were scored correct for turning in the direction of the adult's pointing and head turn and were not required to locate the target object in gaze- and point-following trials. It has been argued that the act of orienting to the same spatial location as adults does not necessarily indicate that joint attention has been established (Tomasello, 1995). In other words, infants as young as 6 months old may not appreciate that others perceptually experience objects. Nevertheless, results of this study suggested that even the earlier behavioral manifestation of gaze following at 6 months may be related to language ability.

In contrast to the previous studies, Watt, Wetherby and Shumway (2006) found in a large (160 children) longitudinal study that attention-following behavior at 12-16 months and at 18-22 months did not predict receptive or expressive language outcome in the third year. A possible explanation for this finding is that their measure of attention-following included only two probes of gaze- and point-following which may have limited the variability on this item (Watt, Wetherby & Shumway, 2006). In addition, unlike all the measures of attention-following used in previous studies, probes used in this study were accompanied by a verbal prompt 'look'. It is worth noting however that attention-following correlated concurrently with a behavioral sample of language comprehension (the ability to show the understanding of different words). Furthermore, it correlated longitudinally with a behavioral sample of language comprehension and language production (inventory of words produced) around 6 months later.

The associations found between responding to joint attention and language in all of the previous studies were assessed using tasks of point-following or gaze-following accompanied by vocalizations such as calling the child's name and head turns which are similar to naturally occurring interactions. It has been suggested that infants understand

the value of the eyes from a very young age and are able to follow gaze without additional linguistic or pointing cues (Brooks & Meltzoff, 2005). Using a cross-sectional design, the authors showed that infants as young as 10 months of age followed adults' head turn significantly more when the adult's eyes were open in comparison to head turns with eyes closed. Furthermore, relations between infant's gaze-following frequency in the open eyes condition and language were analyzed. Results showed that infants' gaze-following and vocalizations at 10 and 11 months were related to their receptive language at 14 and 18 months of age. On the other hand, relations between gaze-following and expressive language were non-significant. The authors argued that infants' understanding that head turns alone are not crucial in monitoring others is an indication of their progress towards acquiring an adult-like understanding of intentional state of others.

Significant relations between expressive language development and gaze-following without additional gestural and vocal cues were also found in a longitudinal study using growth curve analyses of expressive language through two years of age (Brooks & Meltzoff, 2008). However, this relation was reported between average duration of gaze following and not the frequency of gaze following. Furthermore, receptive language was not assessed in this study.

Table 1: Studies examining relations between responding to joint attention and language in typically developing children

Study	Sample	Responding to Joint Attention Task	RJA Scoring Criteria	Language Measures	Key Results
Carpenter et al., 1998	(n = 24) assessed monthly from 9-15 months	No. of targets: 4 Location of targets: left and right Task: **Gazing** (2trials) or **pointing** (2trials) accompanied by calling child's name and showing excited facial expression	Age of Emergence (**AOE**): age the infant first passed the 2 trials of either gaze following or point following Passed if they localized the target object	**Referential** language: spontaneous production of words during session	**Cross-lagged correlations** ● RJA at 9 and 10 months correlated sig. with **referential** language at 12 months (r = .69 and .52, p < .005 respectively) **AOE Correlations** ● RJA **AOE** correlated sig. with **referential** language (r = .36, p < .05)
Beuker et al., 2013	(n = 23) assessed monthly from 8-24 months	No. of targets: 6 Location of targets: within visual field (left/right sides) outside visual field (behind) Task: Examiner **gazed** at the first 3 targets then gazed and **pointed** at the last three targets Maximum of 3 prompts	**AOE**: age the infant first passed the skill: ● Follow **gaze within visual field:** 2 objects ● Follow **gaze outside visual field:** 1 object ● Follow **point within visual field:** 2 objects ● Follow **point outside visual field:** 1 object Passed if they localized the target object	● MCDI: **receptive** and **expressive** vocabulary at 8, 12, 18, 24 months	**Effect size correlations** ● RJA within visual field at 10 and 11 months correlated with **receptive** vocabulary at 18 months (r_{pb} = .52) ● RJA outside the visual field at 14 months correlated sig. with **receptive** vocabulary at 18 months (r_{pb} = .45) **AOE correlations** ● AOE of RJA within and outside the visual field correlated sig. with **receptive** vocabulary at 18 months (r = .42)
Delagado et al., 2002	(n = 47) RJA at 15monhts Language outcome at 24 months	ESCS[a] No. of targets: 4 Location of targets: right, left, right behind and left behind Two sets of **pointing** trials	Scored correct if infant turned to direction of tester's **point** (approx. 45° or 90° off midline for left/right trials or behind trials respectively) **Percent** correct score calculated	● Reynell at 24 months: **receptive** and **expressive** language	**Regression** ● RJA on combination left/right and behind trials sig. predicted **expressive** language (R^2 = .30, p < .01) but not receptive ● RJA on behind trials sig. predicted **expressive** language (B = .07, p < .01) but not receptive when RJA on left/right was controlled for **Partial correlation controlling for cognitive abilities** ● RJA for behind trials correlated sig. with **expressive** language (r = .43, p < .01) but not receptive language

Table continued overleaf

Study	Sample	Responding to Joint Attention Task	RJA Scoring Criteria	Language Measures	Key Results
Morales et al. 2000	(n = 22) RJA every 2 months (6-12 months) then every 3 months (12-24 months) Language outcome at 24 and 30 months	ESCS[a] No. of targets: 3 Location of targets: right, left and behind Two sets of RJA trials ● At 6 months an adapted version of ESCS used. Mothers turned head to target and said child's name 3 times ● At 8-24 months experimenter turned and pointed while saying child's name 3 times	Scored correct if infant's first gaze direction matched adult's direction **Percent** correct score was calculated	● MCDI **expressive** vocabulary at 24 and 30 months ● Direct assessments of **receptive (**PPVT-R) and **expressive** (EVT) vocabulary at 30 months	**Correlations** between RJA and language ● RJA at different ages between 6-18 months sig. correlated with **receptive** and/or **expressive** vocabulary at 24 and/or 30 months **Correlations** between an aggregate score of RJA and language ● RJA 6-18 months aggregate score correlated sig. with MCDI **expressive** (r = .55, p < .05) and direct assessment of **receptive** (r = .66, p < .01) and **expressive** vocabulary (r = .65, p <.01) at 30 months **Regression** (outcome: language at 30 months) controlling for language at 24 months RJA uniquely predicted **receptive** and **expressive** language
Mundy et al., 2007	(n=95) RJA (9, 12, 15 and 18 months) Language outcome at 24 months	ESCS[b] No. of targets: 4 Location of targets: right left, right behind and left behind Two sets of pointing trials	Scored correct if infant's first gaze direction matched adult's direction **Percent** correct score was calculated	● Reynell **receptive** and **expressive** ● MCDI **expressive** ● **Composite:** Reynell (receptive & expressive + MCDI)	**Correlations** ● RJA at 9 months correlated sig. with **receptive** language (r = .35, p < .01) ● RJA at 12 months correlated sig. with **receptive** language (r = .24, p < .05) ● RJA at 9 correlated with MCDI **expressive** (r = .24, p < .05) and the language **composite** (r = .29, p < .05) **Regression** (outcome: **receptive** or the **composite** language score at 24 months) controlling for IJA and cognitive abilities ● RJA at 9 months uniquely predicted **receptive** language ● RJA at 9 months uniquely predicted language **composite**

Table continued overleaf

Study	Sample	Responding to Joint Attention Task	RJA Scoring Criteria	Language Measures	Key Results
Watt, Wetherby, & Shumway, 2006 Longitudinal	(n = 160) children scored ≥ 75 on MSEL at around 3 years CSBS-DP (RJA and language): **Time 1**:12-16 months **Time 2**: 18-22 months Language outcome mean age 33 months	CSBS-DP No. of targets: 2 Location of targets: side and behind the child Examiner gazed and **pointed** while saying "look"	Scored correct if looked where the clinician was **pointing** Scores ranged from 0-2	CSBS DP Behavioral sample: ● Comprehension of simple commands and inventory of words at 12 to 16 months and between 18 to 22 language outcome ● MSEL **receptive** and **expressive**	**Correlations** between RJA and CSBS DP language (concurrent and predictive) controlling for age: ● RJA at 12 to16 months correlated sig. with concurrent **comprehension** (r = .29, p < .001) ● RJA at 12 to16 months correlated with **comprehension** (r = .36, p < .001) and **inventory of words** (r = 35, p < .001) at 16 to 18 months **Correlations with language outcome** at 33 months: ● RJA did **not** correlate with receptive or expressive language
Brooks & Meltzoff, 2005 cross-sectional	(n = 96) 32 infants divided into 3 age groups: 9,10 and 11 months language outcome at 14 and 18 months	No. of targets: 2 Location of targets: right and left 4 trials presented Examiner turned head silently with closed eyes or open eyes towards target for 6 sec	Correct if infant turned and aligned head and eyes with target for at least .33 sec (first look) correct looks: +1, opposite looks: -1, no looking: 0 Looking score: sum of correct looks, incorrect looks, and nonlooks Possible range of scores: -4 to +4	MCDI: **receptive** (words, phrases), **expressive** vocabulary and gestures	**ANOVA effect of condition** ● At 9 months looking score in open eyes condition was not sig. different from closed eyes condition (p > .50) ● At 10 and 11 months looking score in open eyes condition > closed eye condition (p < .05) **Correlation** between looking score (at 10 and 11 months) and language ● Looking score did not correlate sig. with language at 14 or 18 months. ● Looking score correlated sig. only with total gestures at 18 months ● Correct gaze + simultaneous vocalization score at 10-11 months correlated sig. with **receptive** language at 14 months (words: r = .49, p < .05, phrases: r = .57, p < .01) and 18 months (words: r = 64, p < .001, phrases: r = .47, p < .05) and total gestures

Table continued overleaf

Study	Sample	Responding to Joint Attention Task	RJA Scoring Criteria	Language Measures	Key Results
Meltzoff & Brooks, 2008 Longitudinal	(n = 32) RJA at 10 and 11 months Language at (10, 11, 14, 18 and 24 months)	Same as (Brooks & Meltzoff, 2005) but only open eyes condition	Same as (Brook & Meltzoff, 2005) + *average duration* of correct looking score (total duration of correct looking score divided by number correct) and *latency* of correct looking (total latency of correct looks divided number correct)	● CDI words and gestures at 10 months to 1.6 years and CDI words and sentences after 1.6 years **expressive** vocabulary	**Growth curve modelling** ● Frequency or latency measures did not predict accelerated growth ● Average **duration** of correct looking score had a sig. effect ($p < .001$) on the **productive vocabulary growth** model even after accounting for age and maternal education **Chi-square test for vocabulary at age 2** ● More infants from the long duration group had vocabulary scores above the 50th percentile of CDI norms compared to the short duration group

RJA: responding to joint attention; IJA: initiating joint attention; MCDI: MacArthur Communicative Development Inventory (Fenson et al., 1994); ESCS: Early Social Communication Scales ([b]Mundy et al., 2003; [a]Mundy, Hogan & Doehring, 1996); Reynell: Reynell Developmental Language Scales (Reynell & Gruber, 1990); PPVT-R: Peabody Picture Vocabulary Test—Revised (Dunn & Dunn, 1981); EVT: Expressive Vocabulary Test (Williams, 1997); CSBS-DP: Communication and Symbolic Behavioral Scale Developmental Profile (Wetherby & Prizant, 2002); MSEL: Mullen Scales of Early Learning (Mullen, 1995); CDI: MacArthur-Bates Communicative Developmental Inventories (Fenson, Marchman, Thal, Dale, Bates, & Reznick, 2007).

2.3.4 Relations between responding to joint attention and language in autism

A similar picture of relations between RJA and language development that have been reported in the literature of typical development has also been found in clinical children (Sigman & Ruskin, 1999), with most of the research coming from studies of children with ASD (see Table 2 for a summary of studies examining relations between RJA and language in ASD). Unlike studies of typically developing children in which RJA was usually first assessed in the first or early second year of life, studies of children with autism have varied greatly in the age of children at recruitment and assessment of RJA ranging from 15 months to 93 months. This is most likely due to the evidence of delay in the development of RJA in children with autism as compared to children with typical development and other types of developmental delays (Sigman & Ruskin, 1999). Despite this variability in the age of children correlational relations have consistently been reported between RJA and language (Luyster, Kadlec, Carter, & Tager-Flusberg, 2008; Pickard & Ingersoll, 2015; Toth, Munson, Meltzoff & Dawson, 2006; Thurm, Lord, Lee & Newschaffer, 2007). Furthermore, correlational relations were found between RJA in children with autism and later language up to mid-school, adolescents and early adulthood when initial language level was controlled for (Sigman & McGovern, 2005; Sigman & Ruskin, 1999; Siller & Sigman, 2008).

In addition, given the significant variability in the language skills of children with autism, some researchers have examined whether RJA differentiated between groups with good or poor language outcome. Findings suggest that RJA differentiated significantly between outcome groups. For example, Thurm, Lord, Lee and Newschaffer (2007) reported that RJA at age 2 differed significantly between subgroups of children with adequate mental abilities who did or did not acquire language by age 5 when receptive or expressive language status at outcome was examined. Likewise, Paul, Chawarska, Cicchetti, and Volkmar (2008) reported that measures of RJA in young children with ASD (mean age 21.8 months) significantly differentiated groups with good and poor expressive language outcome followed up by age four when nonverbal abilities were already accounted for.

On the other hand, inconsistent findings have been reported with regard to the unique contribution of RJA to language outcome when regressional analyses were used. Most of the studies have used a similar measure of RJA, but there were large differences in the samples and language measures which may explain the inconsistencies in the findings. Furthermore, studies have differed in the variables included in the predictive model of

language outcome. Details of the studies are presented in Table 2. However, a general comparison of the studies is presented in the following section in an attempt to better understand the nature of relation between RJA and language in ASD. Furthermore, considering the findings from studies of both typically developing children and children with ASD will provide more information on the clinical utility of the assessment of RJA in clinically referred children using the ESB.

Two large studies have examined the unique contribution several sociocognitive skills including RJA to language outcome in children with ASD using regressional analyses. Luyster, Kadlec, Carter, and Tager-Flusberg (2008) examined this relation concurrently in a group of 164 children with ASD between 18 and 33 months. Results showed that RJA uniquely predicted concurrent receptive language score when nonverbal cognitive abilities, other social-cognitive skills and motor skills had been accounted for. On the other hand, RJA did not contribute significantly to the model predicting expressive language. In the other study, Thurm, Lord, Lee and Newschaffer (2007) have looked at predictive relations between several sociocognitive skills including RJA and later language in children with ASD. Their results showed that RJA at 2 years of age was a significant predictor of receptive language at 5 years of age. On the other hand, RJA was not a significant predictor of expressive language.

Unlike the findings of these two studies, Toth, Munson, Meltzoff and Dawson (2006)-who also examined the contribution of a number of sociocognitive skills to language ability-reported that RJA in 3 to 4 year old children with autism was not a unique contributor to concurrent receptive or expressive language nor was it a predictor of the rate of communication development between 4 and 6.5 years of age. Instead, initiating protodeclarative joint attention and immediate imitation were predictors of concurrent receptive and expressive language, and toy play and deferred imitation were predictive of rate of development of communication skills. However, children in Toth et al.'s study were older than children in previous studies at intake and had higher language skills which might account for the differences in the results. Thus, as argued by Toth and colleagues, relations between sociocognitive abilities and receptive and expressive language might differ depending on the language stage of the children. RJA might show stronger relations with language when RJA is assessed during the early stages of language acquisition.

Likewise, in a recent study, Van der Paelt, Warreyn, and Roeyers (2014) suggested that RJA relations with language differ depending on the language age. In this study the authors examined relations between different social-communicative abilities and language in a large group of children (83 children) with ASD ranging in age from 22 to 75 months. The sample was divided into subgroups based on their expressive or receptive language level, with language age of 2 years and above or less than 2 years as the dividing criterion. Analyses were done separately for expressive and receptive language. Results showed that RJA correlated concurrently with receptive language only in the children with low language level (i.e. language age less than 2 years). However, unlike Luyster and colleagues' study regressional analyses in this study showed that RJA was not a significant contributor to the variance of concurrent receptive language in either subgroup. Pretend play, imitation, and initiating joint attention showed different unique relations with receptive and/or expressive language depending on the language level of the children. There are a number of possible explanations for the inconsistent findings between Luyster and colleagues' study and Van der Paelt and colleagues' study. First, Van der Paelt and colleagues recruited children from a wide age range and nonverbal cognitive abilities were not controlled for. In addition, Luyster and colleagues used a language composite score formed from three language measures, while language was assessed by a Dutch version of the Reynell Developmental Language Scales in Van der Paelt and colleagues' study.

In contrast to the findings of Van der Paelt and colleagues' study, Pickard and Ingersoll (2015) who also recruited children with autism from a wide age range (22-93 months) found using regression analysis, that RJA was a unique concurrent predictor of a composite measure of receptive and expressive language when chronological age and nonverbal abilities were controlled for. The discrepancy between the results of the two studies might also be attributed to the differences in the variables examined as predictors of language outcome. Unlike Van der Paelt and colleagues who examined different social-communicative skills such as play and imitation as predictors of language outcome, Pickard and Ingersoll focused only on examining relations between the different skills of joint attention and language. In addition, different measures of RJA were used in the two studies.

Table 2: Studies examining relations between responding to joint attention and language in children with ASD

Study	Sample	Language Level	Joint Attention Task	Language Measures	Variables Included in the Regression Models	Key Results
Sigman & Ruskin, 1999 **Longitudinal** (mid-school follow up)	ASD (n = 70) **Time 1**: mean 3;11 years **Time 2 (n = 51):** mean 12;10 years	**Time 1** language age in months: mean 16.6, SD 7.64	ESCS[a]	Time 1: linguist Time 2: Depending on language ability: Reynell or CELF (**receptive** and **expressive**)	N/A	**Correlations** ● **RJA** in children with ASD **correlated concurrently** with overall **language** (r = .72, p < .001), as well as with **receptive** and **expressive** language ● **RJA** in children with ASD **correlated** sig. with expressive language at follow up (r = .44, p < .01) with initial CA and language age partialled out
Sigman & McGovern 2005 **Longitudinal** (adolescence and early adulthood follow up)	ASD (n = 48 of the 70 in Sigman & Ruskin, 1999) **Time 1**: mean 3;11years, SD 1year **Time 2**: mean 12;8 years, SD 3.9 years **Time 3**: mean 19 years, SD 3.10 years	Mid-school language age in months Low IQ < 70 (mean 26.30, SD 9.12) High IQ ≥ 70 (mean 87.04, SD 18.75)	Modified version of ESCS[a]	Depending on language abilities: Reynell or CELF-R or CELF-P (**receptive** and **expressive**)	N/A	**Correlations controlling for initial language** (in preschool) ● RJA in preschool predicted gains in language from preschool to adolescence/young adulthood r (35) = .28, p < .05 ● RJA in mid-school did not predict language gains
Siller & Sigman, 2008 **Longitudinal**	ASD (n = 28) **Time 1**: (mean 45.2 months, SD 8.4, range 31-64) **Time 2**: (mean 57.5 months, SD 8.9) **Time 3**: (mean 69.5 months, SD 9.1) **Time 4**: (mean 89 months, SD 9.1)	**Time 1** language age in months: mean 16.6, SD 7.1	ESCS[a]	Depending on language abilities : ● Reynell ● MSEL[a] ● CELF **(receptive** and **expressive)**	N/A	**Correlation** ● RJA correlated sig. with language at all times of assessment, time 1 (r = .78, p < .001), time 2 (r = .84, p < .001), time 3(r = .85, p < .001), time 4 (r = .86, p < .001) multilevel models ● RJA predicted children's subsequent language gain

Table continued overleaf

Study	Sample	Language Level	Joint Attention Task	Language Measures	Variables Included in the Regression Models	Key Results
Paul, Chawarska, Cicchetti, & Volkmar, 2008 **Longitudinal**	ASD (n = 37) **Time 1**: 15-25 months **Time 2**: 36-58 months	**Time 1** language age in months **:** VABS EL AE (mean 11.4, SD 5.6) VABS RL AE (mean 14.6, SD 6.8) CDI EV AE (mean 13.2, SD 4.1) CDI RV AE (mean 13.6, SD 3.6) **Time 2** Good or poor outcome defined as scores above or below 30 months on VABS EL	CSBS-DP ADOS-1	**Time 1:** ● MSEL[a] **receptive** and **expressive** language ● MCDI words and gestures ● VABS **receptive** and **expressive** **Time 2:** ● **Good or poor language outcome** (VABS EL) ● **Composite** spoken language outcome (from MSEL[a] EL, VABS EL and ADOS communication)	● Nonverbal IQ (MSEL[a] VR) ● Expressive language (VABS) ● Receptive language (VABS) ● Symbolic play (CDI) ● Stereotypic behaviors (ADOS-1)	**Multivariate analysis of covariance** controlling for IQ (MSEL[a] VR) to examine differences between good and poor language outcome groups: ● Outcome groups were sig. different on the RJA (ADOS-1) with large effect size **Regression** (outcome: **composite expressive score**): ● RJA did not make sig. contribution to the model, only receptive language and stereotypic behaviors were sig. contributors to language outcome
Thurm, Lord, Lee & Newschaffer, 2007 **Longitudinal**	ASD (n = 110) **Time 1**: 2 years **Time 2**: 5 years analysis reported for a subsample with no receptive or expressive language at 2 years and IQ >18 months at 5 years	**Time 1** *no language* (expressive or receptive) defined as follows: *Expressive* (<5words or speech not used on a daily basis on ADI-R overall level of language question) *Receptive* (≤ 18 months on a language test)	PL-ADOS	Measures were used to categorize children's language **Time 1:** ● ADI-R: overall level of language ● DAS **Time 2:** ● ADI-R ● DAS ● MSEL[b,c]	● IJA (PL-ADOS) ● imitating sounds (SICD) ● imitating simple movements (VABS)	**T-test**: Groups with language (receptive or expressive) or no language (receptive or expressive) at 5 years were sig. different on RJA **Regression** (outcome**: receptive or expressive language**): ● RJA sig. predictor of **receptive** language ● RJA not a sig. predictor of expressive language, only imitating sounds sig. predicted expressive language

Table continued overleaf

Study	Sample	Language Level	Joint Attention Task	Language Measures	Variables Included in the Regression Models	Key Results
Luyster, Kadlec, Carter, & Tager-Flusberg, 2008 **Concurrent**	ASD (n = 164) Age: 18-33 months	Language age in months: Receptive (MSEL[a]: mean 16.25, SD 6.35), (VABS: mean 20, SD 7.91) Expressive (MSEL[a]: mean 16.75, SD 7.49), (VABS: mean 31.33, SD 16.30)	ESCS-L	**Composites** for receptive and expressive from (MSEL[a] + VABS communication + MCDI)	● Nonverbal IQ (MSEL VR) ● IJA (ESCS-L) ● Imitation (IB) ● Gestures (MCDI) ● Play: functional and symbolic (ADOS-G) ● Motor (MSEL[a] fine and gross motor, VABS fine and gross motor)	**Correlation** controlling for age: ● RJA **correlated** sig. with **receptive** (r = .55, p < .001) and expressive (r = .57, p < .001) composites **Regression** controlling for age: ● **RJA** was a **sig. concurrent** predictor of **receptive** language as well as gestures and nonverbal cognitive ability ● RJA was not a sig. predictor of expressive language. Imitation, gestures and IQ were sig. predictors of **expressive** language
Toth, Munson, Meltzoff & Dawson, 2006 **Longitudinal**	ASD (n = 60) **Time 1:** 34-52 months **Time 2:** 65-78 months	**Time 1** MSEL[d] Verbal AE=average of expressive AE and receptive AE in months: mean 22.9, SD 10.3, range 8-50	ESCS[a]	● MSEL[d] ● VABS overall communication subscale (receptive, expressive and written communication) was used in growth curve analysis	● IJA (ESCS) ● Imitation: immediate & deferred (Meltzoff, 1988a,b) ● Toy play functional and symbolic	**Concurrent correlation** ● RJA **correlated** sig. with all measures of language (MSEL[d] **receptive** and **expressive** and VABS communication) **Regression** (outcome: MSEL[d] verbal AE, MSEL[d] RL AE, MSEL[d] EL AE, or VABS) ● RJA did not predict concurrent language. Initiating protodeclerative joint attention and immediate imitation sig. predicted concurrent language. **Hierarchical linear modeling** predicting communication abilities at 48 months on VABS ● Immediate imitation and toy play abilities were sig. related to individual differences in children's communication ability at 48 months ● Toy play and deferred imitation were sig. related to *rate* of acquisition of communication skills
Van der Paelt, Warreyn, & Roeyers, 2014 DUTCH **Concurrent**	ASD (n = 83) Age: 22-75 months	Children grouped based on language level above or below 2 years for receptive and expressive	ESCS[a]	Reynell-Dutch version: **receptive** and e**xpressive**	● Imitation (PIPS) ● Pretend play (ToPP) ● IJA IBR (ESCS)	**Correlation:** ● **RJA** correlated sig. with **receptive** language **ONLY** in **children with language level < 2 years of age** (r = .41, p<.01). **Regression:** RJA did not contribute sig. to receptive or expressive language ● Pretend play and IJA were sig. predictors of variance in receptive language < 2 years ● Imitation and IBR were sig. predictors of expressive language <2 years ● Pretend play was a sig. predictor of receptive language > 2 years ● Imitation and pretend play were sig. predictors of expressive > 2 years

Table continued overleaf

Study	Sample	Language Level	Joint Attention Task	Language Measures	Variables Included in the Regression Models	Key Results
Pickard and Ingersoll, 2015 **Concurrent**	ASD (n = 53) Age: 22-93 months	Language age in months: Receptive language AE; PLS-4/MSEL[a] (mean 18.57, SD 10.08, range 4–46) Expressive language AE; PLS-4/MSEL[a] (mean 20.57, SD 8.54 range 6–43) Expressive vocabulary MCDI no. of words (mean 157.59, SD 181.22, range 0–628)	ESCS[b]	● PLS-4 or MSEL[a] **receptive** and **expressive** ● MCDI **expressive vocabulary** **composite language** (receptive, expressive, vocabulary)	● IQ: BSID-III or MSEL ● IJA (ESCS) ● Imitation: (MIS, UIA-O)	**Correlation:** ● RJA **correlated** with **receptive** language (r = .73, p < .002), **expressive language** (r = .65, p < .002) and **expressive vocabulary** (r = .64, p < .002) **Regression (**outcome: language composite) controlling for CA and IQ: ● **RJA** uniquely **predicted** composite language

------ Follow up study; RJA: responding to joint attention; ASD: autism spectrum disorder; SD: standard deviation; ESCS: Early Social Communication Scales ([a]Mundy, Delgado, Hogan & Doehring, 2003; [b]Siebert et al., 1982); Reynell: Reynell Developmental Language Scales (Reynell & Gruber, 1990); CELF-R: Clinical Evaluation of Language Fundamentals-Revised (Semel, Wiig, & Secord, 1987); CELF-P: Clinical Evaluation of Language Fundamentals-Preschool (Wiig, Secord, & Semel, E, 1992); VABS: Vineland Adaptive Behavioral Scales (Sparrow, Balla & Cicchetti, 1984); EL: expressive language; AE: age equivalent; RL: receptive language; CDI: MacArthur Communicative Development Inventory—Words & Gestures (Fensen et al., 2002); EV: expressive vocabulary; RV: receptive vocabulary; VR: visual reception; CSBS-DP: Communication and Symbolic Behavioral Scale-Developmental Profile (Wetherby & Prizant, 2003); ADOS-1: Autism Diagnostic Observation Scale-General; Module 1 (Lord et al., 2000); ADI-R: The Autism Diagnostic Interview-Revised (Lord, Rutter, & LeCouteur, 1994); DAS: Differential Ability Scale (Elliot, 1990); MSEL: Mullen Scales of Early Learning (Mullen, [c]1985, [b]1989, [a]1995, [d]1997); PL-ADOS: Pre-Linguistic Autism Diagnostic Observation Schedule (DiLavore, Lord, & Rutter, 1995); SICD: Sequenced Inventory of Communication Development (Hedrick, Prather, & Tobin, 1975); ESCS-L: Early Social Communication Scales-Live scoring (Thorp & Mundy, 2010); MCDI: MacArthur-Bates Communicative Development Inventory (Fenson et al., 1993); IJA: initiating joint attention; IB: Imitation Battery (Rogers, Hepburn, Stackhouse & Wehner, 2003); Reynell-Dutch version: Reynell Developmental Language Scales-Dutch Version (Schaerlaekens, Zink, & Van Ommeslaeghe, 2003); PIPS: Preschool Imitation and Praxis Scale (Vanvuchelen, Roeyers, & De Weerdt, 2011); ToPP: Test of Pretend Play (Lewis & Boucher, 1997); PLS-4: Preschool Language Scales, 4th Edition (Zimmerman et al. 2002); BSID -III: Bayley Scales of Infant Development-3rd Edition (Bayley, 2006); MIS: Motor Imitation Scale (Stone et al., 1997); UIA-O: Unstructured Imitation Assessment-Object Scale (Ingersoll & Meyer, 2011).

2.3.5 Joint attention and language conclusion

There is replicated evidence from studies of typically developing children and studies of children with autism of concurrent and predictive relations between RJA and language. Interestingly, when relations with both receptive and expressive language outcome were examined, significant relations were mostly reported with receptive language despite the large differences in methods of the studies. These findings support the sociocognitive hypothesis which suggests that impairments in responding to joint attention will impact children's ability in inferring the speaker's referential intent and determining meanings behind people's words (Chiat & Roy, 2013). In addition, though studies are not comparable in several aspects such as participants' age, diagnosis, and measures of RJA or language, a number of key findings can be drawn from the reviewed literature which will have implications on the clinical utility of assessing RJA using the ESB.

1. Studies of typically developing children showed unique predictive relations between RJA and language when RJA was assessed in the first or early in the second year of life. This is likely the age in which variability in performance on tasks of RJA may be observed. By 18 months, most typically developing infants show capability of following attention on most trials of RJA which limits variability on performance on this measure.

2. Studies of children with autism showed that unique predictive relations between RJA and language in children with autism were more likely to be found among children in the early stages of language acquisition when other skills of sociocognition had been accounted for.

3. In children with autism who began using language, other sociocognitive skills such as play or imitation were suggested to be more important to language outcome. However, this does not mean that RJA at this stage is not related to language outcome in autism as correlational relations have been reported between RJA with language outcome in studies that greatly varied in age and language skills of participants. Unlike typically developing children, variability in RJA abilities may be observed between older children with autism.

Collectively these findings suggest that responding to joint attention is an important aspect to consider in the assessment of young children with concerns about language particularly in the early stages of language acquisition. Accordingly, the inclusion of a measure of RJA in the ESB contributes to the picture of the child's sociocognitive abilities that are

important to the acquisition and use of language. Furthermore, this measure considered several issues that are important when examining joint attention that have not been necessarily considered in other assessments and may make it more sensitive to the variability of RJA. For example, most measures of joint attention operationally defined joint attention as the ability to follow the examiner's gaze and or point. It has been argued that this behavior alone is not indicative of the child's ability to read intentions and only when the behavior of gaze and point following are accompanied by checking behaviors or gaze switching that it seems to indicate the child's understanding of others as intentional agents (Carpenter et al., 1998). RJA as measured on the ESB includes a measure of gaze switch and a measure of gaze or point following. Another issue that has been considered in the scoring of RJA in the ESB is differentiating between the scores accredited when the child follows the gaze versus when the child follows the point. This issue has not been considered in most measures whereby children received the same score whether they followed the adult's point or gaze. Furthermore, many measures used only a point-following task. Thus, it is assumed that this measure will be a relatively sensitive measure of RJA and informative of concurrent and later language of children in the targeted age range.

To conclude, this section has shown that the ability to follow and understand the focus of the social partner facilitates language acquisition. During these episodes of joint attention infants may gain access to the intention of the speaker's words. However, in order to understand and use words in different contexts it is also critical that children have symbolic abilities (see below).

2.4 Symbolic Comprehension and Language

A symbol is defined as "something that someone intends to stand for or represent something else" (DeLoache, 2002, p.73). Thus, DeLoache emphasizes that human intention is fundamental for the establishment of symbolic relations. Symbol formation has been proposed to be an important prerequisite for language (Bates, 1979). In order to use words children have to understand that they are linguistic symbols standing for referents of speaker's intended meaning. Accordingly, many researchers have investigated the relations between different forms of symbolic ability and language such as the understanding and use of gestures and symbolic play. The general consensus has been that

early language and different forms of nonverbal symbolization are closely related developmentally. For example, with regard to symbolic or pretend play - which involves either the attribution of absent properties to an object, substitution of one object for another or the imagination of absent objects (Leslie, 1978) - close developmental relations have been repeatedly found with language. In an observational study of typically developing children between the ages of 8 and 24 months, McCune (1995) reported that language and play are tightly coupled in development. Significant relations were found between the onset of symbolic play and beginning of lexical development and between sequences of symbolic play and the onset of word combinations. In addition, ten participants were followed longitudinally from 8 or 10 months up to 24 months. Analyses showed that new levels of language skills emerged in most of the children 2 months after the emergence of the proposed equivalent symbolic play level. The authors proposed that this relationship is due to a common underlying representational system.

Relations between pretend play and language were also documented in special populations including children who have hearing impairment (Spencer, 1996), autism (Toth, Munson, Meltzoff & Dawson, 2006) and Down's syndrome (O'Toole & Chiat, 2006). Researchers have repeatedly found concurrent relations between different measures of symbolic play and different aspects of language. In addition, it was shown that symbolic abilities are predictive of later language outcome (Toth, Munson, Meltzoff & Dawson, 2006).

Furthermore, it was reported that toddlers with expressive language impairment differed from a comparison group of age matched typically developing children in object based symbolic play (Rescorla & Goossens, 1992). Two-year olds in this study where observed during free play and structured play sessions. It was found that 'late talkers' produced less advanced object based symbolic play such as using substitute or imaginary objects. Furthermore, their play sequences tended to be shorter and involving smaller range of play behaviors in comparison to typically developing children.

A number of researchers, however, have questioned the extent to which children's early pretense reflects their true symbolization ability (Tomasello, Striano, & Rochat, 1999). The production of symbolic play is often assessed through natural observations which involve a parent interacting with the child, which means that the parent may scaffold their symbolic play abilities through language and modeling (Tomasello, et al., 1999). Alternatively, symbolic play is assessed using structured methods of assessment in which

the child might be verbally instructed to perform certain acts or asked to imitate a model provided by the examiner. Thus, the child may be simply reproducing the use of the symbol and this might not reflect their true representational ability (Casby, 1997). Furthermore, some researchers have questioned whether examining the child's production of symbolic play during structured assessment reflects their representational ability as opposed to performance ability which may be affected by the child's ability to generate ideas for pretense (Bigham, 2008). In addition, children's production of symbolic play is probably affected by their willingness to engage with a strange adult (the examiner) in a playful manner.

Accordingly, using structured symbolic comprehension tasks that require the child to respond in a simple way and are not heavily dependent on language or modeling may provide clearer evidence of the child's representational ability. This approach was used in a task developed by Tomasello and colleagues (1999) and was further adapted by Chiat and Roy (2006a) in the symbolic comprehension task in the ESB.

While the extent to which pretend play truly reflects representational ability is a controversial issue. The observed correlations between pretend play and language suggest some common underlying functions (Lillard, 2011). In addition, in some studies in which both comprehension and production of symbols were examined correlations between the two skills were found suggesting that they measure representational skills in a similar way (O'Toole & Chiat, 2006). Though, this relationship may defer depending on the tasks used or the level of symbolic ability being assessed. For example, Lillard and Kavanaugh (2014) reported no correlations between production of symbolic play when measured in a free play and a task of comprehension of pictures used as symbols.

Interestingly, while language comprehension precedes production, the opposite pattern was found with regard to pretend play, with pretend play production emerging before comprehension (Lillard & Kavanaugh, 2014). In pretense comprehension the child reads the symbols created by others. This ability has been suggested to depend on several sociocognitive skills such as joint attention, social referencing and sensitivity to other people's intentions (Lillard, 2007, 2011).

Studies that focused on symbolic comprehension showed that not until 2 years of age do children show understanding of symbolic play. For example, Rakoczy, Tomasello and Striano (2004) and Rakoczy and Tomasello (2006) have shown that 2 years old in their

study where able to differentiate between pretending to do an action and trying to do an action. This understanding was reflected in the children's actions and spontaneous language. For example, in the pretend condition children who observed the examiner performing a pretend action showed more related pretense actions when given a chance. On the other hand, in the trying condition children tried to achieve the examiner's goal using other tools or verbally commented on that goal. In order to achieve that understanding children had to have the ability to recognize the signals that mark pretense acts and playfulness such as sound effects, smiling or laughing (Rakoczy & Tomasello, 2006).

This finding has been replicated by other researchers using different methodologies (MacConnell & Daehler, 2004). Furthermore, it was suggested that the understanding of some types of symbols is easier than others. More specifically, the interpretation of gestures used as symbols was reported to emerge before the understanding of objects used as symbols (Tomasello, Striano, & Rochat, 1999). The difficulty of understanding objects used as symbols was suggested to be as a result of "dual representation problem" (DeLoache, 1987, 1989, 1991). That is, young children have a difficulty in understanding that an object is both a concrete object and a representation of something else. Moreover, the interpretation of objects used as symbols becomes more difficult when the object have another conventional use (e.g. a toothbrush as a pen). This has been termed "triune representation" (Tomasello, Striano, & Rochat, 1999). In this case the child has to consider (1) the object's physical and perceptual properties, (2) the object's conventional use, and (3) that the object is being used symbolically in this instance to represent something else.

Like production of symbolic play understanding of symbols have been shown to relate to language. For example, Lillard and Kavanaugh (2014) have shown that language correlated concurrently with different tasks of symbolic understanding including: pictures, miniature objects and the understanding of pretend play that involved attribution of pretend properties (i.e., pretending that the toy is wet when it is not) in children aged 24-36 months.

Likewise, O'Toole and Chiat (2006) examined relations between symbolic comprehension and language in children with Down's syndrome aged 2-7 years old. Their measure of symbolic comprehension is the same measure included in the ESB in which the

understanding of three types of symbols is assessed: gestures, miniatures and substitute objects. Results showed that the total score of symbolic comprehension correlated concurrently with receptive and expressive language when age was partialled out. Furthermore, the authors showed that relations changed over age and only in the youngest age group which is between 2 and 3 years old were correlations significant with language. A similar pattern of relations between symbolic play and language across age groups was reported in an earlier study with typically developing children, although significant correlations between language and symbolic play were found up to 4;9 years old (Doswell, Lewis, Sylva, & Boucher, 1994). The lack of association between symbolic comprehension and language in older age groups in O'Toole and Chiat's study may be explained by comparing improvement with age in symbolic comprehension and language skills. While symbolic comprehension appeared to increase with age, standard language scores tended to decrease with age. Accordingly, the authors concluded that language deficits in children with Down's syndrome cannot be attributed uniquely to their symbolic functioning.

More recently, Maljaars, Noens, Scholte and Berckelaer-Onnes (2012) investigated relations between comprehension of symbols and language in three groups of children: typically developing children (TD), children with autism and children with intellectual disability (ID). The three groups were matched on their nonverbal mental abilities but the group of children with autism had the lowest verbal age. Symbol understanding was measured using a task that involved understanding of pictures used as symbols to represent a physical space whereby children had to retrieve different items when shown photographs of the location of the hidden item in a doll house. Relations between symbolic understanding and receptive or expressive language were examined for each group separately. Thus, a total of 6 regressional analyses were conducted with children's nonverbal mental age included as a covariate in the analyses. Results showed that symbolic understanding was a significant concurrent predictor of receptive and expressive language in children with autism. In addition, symbolic understanding was a significant predictor of receptive language in children with ID. On the other hand, symbolic understanding was not a significant predictor of language outcome in typically developing children. These findings may be explained by the fact that TD children had a higher verbal age. Previous studies showed that relations between symbolic understanding are more pronounced in younger children (Doswell, Lewis, Sylva, & Boucher, 1994), and with

O'Toole and Chiat's conclusion that as children get older and their language skills become more advanced language starts to dissociate from other cognitive and representational skills and become more domain-specific.

In addition, an earlier study by Bigham (2008) suggested that children with autism have a specific difficulty in understanding pretense when compared to language matched controls. In this study, different levels of symbolic comprehension were compared in children with autism, typically developing children and children with moderate learning difficulties matched on receptive language age. Children with autism performed below comparison groups and differences were more notable in understanding pretense that depended on triune representation. The authors concluded that these difficulties are possibly due to inhibitory control difficulties in children with autism. However, it is also possible that the understanding of triune representation depends more on the child's ability to read others' intentions which might be impaired in children with autism. While the interpretation of some symbols may be achieved by relying on other cues such as the physical similarities between the substitute and its referent (Bigham, 2008) interpretation of triune representation depends highly on the child's ability to attend to the examiner's behavior and recognize playfulness cues.

2.4.1 Symbolic comprehension and language conclusion

In sum, nonverbal symbolic abilities appear to be tightly linked to language development during the early stages. Accordingly, the symbolic task in the ESB which measures three different developmental levels of nonverbal symbolization may provide means to understand whether the child's language impairments are due to underlying deficits in symbol formation. This is particularly important with respect to the assessment and intervention of young children with language impairments.

2.5 Conclusion

Collectively, this chapter has shown that the skills assessed in the ESB are key precursors to language development and important predictors of language outcome in both typically developing children and children with ASD. The three tasks span a range of sociocognitive skills from the basic abilities of attending to facial expressions and

following the focus of a social partner to the understanding that different forms can be used as symbols to communicate meaning. Assessing these skills systematically using the ESB in which a certain number of scenarios are set up to elicit certain responses overcomes the problems faced when measuring skills through observations of natural interactions with a parent or when using measures that are highly dependent on child initiations of behavior with an examiner. These problems may include for example the influence of the parent's language on the child's behavior, or the child's willingness to interact with the examiner. Thus, the ESB may provide a more efficient means of assessing these necessary skills. However, these skills, though crucial, are not sufficient for language acquisition. Imitation is another sociocognitive skill that has been shown to have a unique predictive value to language outcome (Toth, Munson, Meltzoff & Dawson, 2006; Van der Paelt, Warreyn, & Roeyers, 2014). Assessing imitation may contribute to the assessment and intervention of children with language impairments.

CHAPTER 3

NONVERBAL IMITATION

3.1 Introduction

Imitation refers to the voluntary reproduction of a behavior modeled by another individual (Butterworth, 1999). It is among the cluster of early sociocognitive skills that have been described as fundamental to the acquisition of language (Carpenter, Nagell & Tomasello, 1998). As reported in the previous chapter, relations have been found between imitation and other social cognitive skills such as joint attention (Pickard & Ingersoll, 2015), play (Stone, Ousley, & Littleford, 1997; Stone & Yoder, 2001), and gestures (Carpenter et al., 1998; Carpenter, Pennington, & Rogers, 2002).

The role of nonverbal imitation in language and communication development in typically developing children and children with autism, and whether performance on nonverbal imitation tasks differentiates children with autism from other children, have been the focus of much research. As reported in the previous chapter, unique relations have been found between nonverbal imitation and later language in children with autism when several other sociocognitive skills had been accounted for (Luyster, Kadlec, Carter, & Tager-Flusberg, 2008; Toth, Munson, Meltzoff, & Dawson, 2006; Van der Paelt, Warreyn, & Roeyers, 2014). Accordingly, the evidence supports the inclusion of a measure of nonverbal imitation as a predictor of language development in children with autism.

However, far less attention has been paid to nonverbal imitation in children with language delay and the potential of using a nonverbal imitation task as a clinical marker for longer term language and communication difficulties in young children with concerns about language. Only a few studies have investigated the nonverbal imitation ability of preschool children with language delays. The evidence discussed in this chapter motivated the inclusion of a motor imitation assessment in addition to the skills assessed in the ESB. At a practical level, identification of children's nonverbal imitation abilities may not only enable clinicians to better predict outcomes of children with early language delay but also enable children to receive more targeted intervention in the skills that are important for language acquisition.

This chapter will briefly explore the role of imitation early in life; examine the evidence of relations between imitation skills and language; and present and discuss findings of studies

investigating imitation performance in clinical groups. Finally, results will be summarized and conclusions will be drawn.

3.2 The Role of Imitation in Early Development

Early in life imitation plays an important role in the development of cognitive and social abilities. Children use imitation to acquire new skills and knowledge. Furthermore, through reciprocal imitation children engage in social and emotional exchanges with others. These intimate interactions provide a sense of connectedness (Uzgiris, 1981).

Piaget (1962) was the first to emphasize the importance of imitation, giving it a central role in the development of language. He described body, vocal and facial imitation development through the six stages of sensorimotor development. Piaget excluded the possibility of neonatal imitation and argued that the infant's early matching behaviors are coordinated reflex-like behaviors which create circular reactions. He linked changes in imitation during the first 8 months to the accommodation function, which refers to the modification of a subject's scheme of actions by the external world, and proposed that imitation of movements that babies cannot see themselves make such as facial gestures begin at stage 4 (8-12 months). Starting from 18 months of age (stage 6), infants are able to reproduce actions they observed after a delay (deferred imitation). From a cognitive perspective, deferred imitation serves as an index to the infant's internal representational capacity (Meltzoff, 1988) which has been linked to the development of language and symbolic play (Piaget, 1962).

In the 1970s and 1980s a series of studies challenged Piaget's account and showed that newborns as young as 42 minutes could imitate various gestures such as mouth and tongue movements (Meltzoff & Moore, 1983, 1989). Meltzoff and Moore (1977) argued that infants' imitations are not general arousal but specific matching behavior in terms of the action used and the body part moved. In a number of studies it was demonstrated that infants vary their responses to match the act shown (e.g. lip protrusion versus mouth opening), and the body part moved (e.g. tongue protrusion versus lip protrusion) (Meltzoff & Moore, 1977).

In Meltzoff's view (2007, 2011), imitation allows infants to perceive others as 'like me'. Through imitation they are able to map the similarities and differences between self and

others which lead to an understanding of others' behavior and form the basis of theory of mind and social cognition. Such a viewpoint is in line with Carpenter and colleagues' (1998) argument that imitation is one of the early skills that leads to the understanding of other persons as intentional beings.

Meltzoff (2005) stated "*If infants can recognize when an entity is acting "like me," this would allow them to make a distinction between people and all other entities in the world"* (p.60). Interestingly, Nadel (2006) showed that 2-month olds infants are selective in their imitation. Infants were presented with a stranger protruding her tongue and a robotic tongue protrusion in counterbalanced order. Results showed a significant difference between the two conditions with infants imitating the human and not the robot. Nadel (2002) suggested that imitation's interpersonal function forms the building blocks for verbal language. According to Nadel, imitation is a tool through which children practice turn-taking, learn to understand others' intentions and share goals. "*The imitative language can therefore be seen as a semantic foundation for verbal language*" (p.58).

Further support for the social role of imitation comes from naturalistic imitation intervention studies in which collateral improvements were noted in other social communicative behaviors including joint attention, pretend play, language (Ingersoll & Schreibman, 2006) and spontaneous gesture use (Ingersoll, Lewis, & Kroman, 2007). Furthermore, intervention effects were generalized to new settings and maintained several months post treatment (Ingersoll, 2012).

3.3 Relations between Imitation and Language

Relations between imitation and language have been established in a vast number of studies using different methodologies in both typically developing children and children with autism. For example, relations were reported when imitation was assessed using parental reports (McEwen et al., 2007), through analyses of spontaneous imitation in home video (Poon, Watson, Baranek, & Poe, 2012) and in structured tasks of elicited imitation (Young et al., 2011). However, researchers have attributed these relations to different underlying skills that may be required for the development of both abilities such as intentional understanding (Carpenter et al., 1998), responsive interaction with others (Sigman & Ungerer, 1984) and representational requirements (Stone et al., 1997). This section will provide an overview of relations between language and motor imitation in

both typically developing children and children with autism. This review focuses on studies that used measures of elicited immediate motor imitation (of hands and arms) in children as the goal is to develop a quick structured test to be used with young clinically referred children. Imitation of oral facial movements, though important, may be linked more to oral motor abilities (Page & Boucher, 1998) and will not be included as the aim of the current study was to examine skills that draw more on sociocognitive abilities.

Researchers have tended to use two main kinds of motor imitation tasks: imitation of body movements and imitation of actions with objects. However, an array of terms has been used in the literature to refer to the various tasks. For the purpose of comparing findings in a coherent way the following terms will be used in the current study. Imitation of body movements refers to imitation acts that do not involve an object or in which a placeholder – a wooden or plastic block – is used for a missing object. This includes imitation of *gestures* which are movements that carry meaning in their form to symbolize a referent (Capone & McGregor, 2004) or *postures* which are non-meaningful gestures that do not have semantic associations or conventional communicative meaning.

Imitation of actions on objects refers to imitation of either meaningful or non-meaningful actions on objects. *Meaningful actions* include familiar actions on *appropriate objects* such as drinking from a toy cup or novel *instrumental* actions with an obvious goal such as manipulating a novel object in a certain way to produce a beeping sound. *Non-meaningful* actions include actions with *inappropriate objects* such as walking a hairbrush across the table or novel *arbitrary* actions achieving an obvious goal in unusual way such as pressing an object with the forehead to turn on a light.

3.3.1 Motor imitation and language in typical development

A number of early studies have examined relations between motor imitation and early language in children in their first year or early second year of life. Concurrent or predictive associations between motor imitation and language have generally been reported. However, the nature of this relationship is not yet clear since most of these relations were examined using correlational analyses. Furthermore, comparability of these studies is limited due to the large differences in the methodologies used in the different studies. These studies differed in the ages of children included, measures of language used

and more importantly in the way imitation was operationally defined, tasks administered, and scoring criteria adopted. Nevertheless, pulling together findings from studies of relations between imitation and language in typically developing children and children with ASD is important to improve our understanding of the nature of relations between imitation and language.

One of the earliest studies that examined relations between imitation and language was by Bates, Bretherton, Snyder, Shore and Volterra (1980). Thirteen month old infants were tested on an imitation task that included imitation of meaningful and non-meaningful actions on appropriate and inappropriate objects. Children's expressive and receptive vocabulary was determined based on parental interviews. Correlational analyses showed that both types of imitation correlated significantly only with children's expressive vocabulary, though correlations were slightly higher between expressive vocabulary and imitation of non-meaningful actions with objects (r = .51, p = .005) compared to imitation of meaningful actions with objects (r =.36, p < .05).

Bates, Thal, Whitesell, Fenson, and Oakes (1989) also examined relations between imitation and language in a sample of typically developing children (13 to 15 months old) but using a gestural imitation task. Children's language was assessed using parental reports of receptive and expressive vocabulary. Like Bates and colleagues (1980), they found that children in the high expressive vocabulary group produced significantly more gestural imitation. On the other hand, relations between receptive vocabulary and gestural imitation were suggested to reflect a U-shaped relation with children in the middle comprehension group scoring the highest on the imitation task.

In contrast, using direct rather than parental assessments, Sigman and Ungerer (1984) found different relations between imitation of body movements (postures and gestures), receptive and expressive language in an older group of children (16-25 months). Children's receptive vocabulary was assessed using a picture pointing task, while expressive language was assessed by a linguist during a semi-structured play task. The authors reported large concurrent correlations between imitation of body movements and receptive vocabulary.

Longitudinal relations between imitation and expressive language have also been found. As reported in the previous chapter, Carpenter, Nagell and Tomasello (1998) examined relations between the age of emergence of a number of social cognitive skills including

following attention, directing attention, communicative gestures, imitative learning (imitation of arbitrary and instrumental actions), and referential language in infants from 9-15 months. Receptive language was not assessed in this study. Results showed a moderate correlation between the age of emergence of imitation of arbitrary actions on objects and referential language. On the other hand, no significant correlations were found between instrumental imitation and referential language.

In contrast, Charman and colleagues (2000) reported no concurrent relations between the imitation of arbitrary and instrumental actions and receptive or expressive language abilities at 20 months of age. However, imitation was predictive only of expressive language outcome at 44 months, although this association was not significant after controlling for initial language ability and IQ. As noted by the authors, this study was limited by the small number of participants, ceiling effect on the imitation task, and the reliability of the language measure at time 1.

To summarize, the large differences between methodologies adopted in the previous studies cautions against simple comparisons. However, these studies suggest that different types of imitation may have different relations with language at different points of development. Furthermore, when examined in the same study, imitations of non-meaningful actions showed stronger relations with language compared to meaningful actions (Bates et al, 1980; Carpenter et al., 1998). Motor imitation relations were consistently reported with children's vocabulary, whether it was receptive (Bates et. al, 1989; Sigman & Ungerer, 1984), or expressive vocabulary (Bates et al., 1980, 1989; Carpenter et al., 1998). In addition, these relations were reported when vocabulary was assessed via parental reports (Bates et al., 1989), parental interviews (Bates et al., 1980), direct assessment (Sigman & Ungerer, 1984) or observation of spontaneous production (Carpenter et al., 1998).

3.3.2 Motor imitation and language in ASD

The association between motor imitation and language has been the focus of much research on children with autism. There is replicated evidence of concurrent and longitudinal correlations between imitation and receptive and expressive language (Ingersoll & Meyer, 2011; Sigman & Ungerer, 1984; Stone, Ousley, & Littleford, 1997). Of particular interest are the unique predictive relations reported between motor imitation

and language with several other sociocognitive skills controlled for. As seen in the previous chapter, this was reported in a number of studies in which different sets of skills were examined as predictors. For example, Toth, Munson, Meltzoff and Dawson (2006) reported that imitation of instrumental and arbitrary actions on objects emerged as a unique predictor of concurrent receptive and expressive language in 3- to 4-year-old children with ASD when several other sociocognitive skills were controlled for, though children's IQ was not controlled for. Furthermore, using growth curve analyses this study showed that immediate imitation of actions on objects uniquely predicted individual differences in communication ability at 48 months. But it was not predictive of rate of development of communication skills between 4 and 6;5 years old, with only toy play and deferred imitation proving predictive of rate of development of communication skills over this period.

On the other hand, imitation emerged as a unique predictor of later language in an earlier study with younger children with ASD (Stone & Yoder, 2001). In this study, the authors examined the predictiveness of different socio-communicative and environmental factors including imitation, joint attention, play, socioeconomic status and speech therapy for the development of expressive language in 2-year-old children with ASD. Children's imitation ability was measured using the Motor Imitation Scale (MIS; Stone, Ousley, &Littleford, 1997) and expressive language was assessed using parental report of expressive vocabulary and direct observation measures. The MIS includes 16 items. Half of the items consist of imitation of body movements (gestures and postures). The other half comprise imitation of actions with objects (meaningful and non-meaningful). The authors reported that among the children's socio-communicative skills only motor imitation at age 2 was a unique predictor of expressive language at age 4 after controlling for initial language level.

Other researchers have questioned whether or not a certain type of imitation showed stronger relations with language. For example, McDuffie, Yoder and Stone (2005) investigated the predictive association of four prelinguistic behaviors at 2 and 3 years (imitation, attention following, commenting and requesting) with later receptive and expressive vocabulary. Imitation ability was assessed using the MIS, but only the sub-scores were used in the analysis: MIS (with objects) and MIS (body imitation). Commenting and motor imitation of body movements were unique predictors of vocabulary production six months later after controlling for the degree of cognitive delay.

Likewise, Stone, Ousley, and Littleford (1997) reported distinct relations between imitation of body movements and language in children with ASD aged 23-35 months who were followed up on average 14 months later, though associations were examined using correlational analyses. Using the subscores of the MIS, they reported that imitation of body movements (gestures and postures) correlated with expressive vocabulary concurrently and predictively, whereas imitation of actions on objects (meaningful and non-meaningful) was related to play skills. The authors attributed the stronger relation between imitation of body movements and language to the higher level of representation in that type of imitation.

More recently, Ingersoll and Meyer (2011) partly replicated the findings of Stone and colleagues (1997) using the same measure of imitation. Imitation of actions on objects showed unique correlations with symbolic play when cognitive ability was partialled out. Furthermore, body imitation correlated with expressive language and expressive vocabulary concurrently. But this relationship was not unique to body imitation; imitation of actions on objects and total imitation also showed concurrent relations to expressive language and expressive vocabulary. However, after controlling for cognitive level, only the total imitation score showed significant correlations with expressive vocabulary. Thus, these results are inconsistent with Stone and colleagues' (1997) findings of a unique relationship between body imitation and expressive vocabulary, despite the fact that both studies used the same measure to assess expressive vocabulary and imitation. It is possible that the discrepant findings were due to differences between the samples in the two studies. Participants in Ingersoll and Meyer's study included a slightly wider range of participants in terms of age and verbal abilities (CA: 22-47 months, words produced: 0-347 words) compared to participants in Stone and colleagues' study (CA: 26-36 months, words produced: 0-150 words).

Carpenter, Pennington and Rogers (2002) examined relations between four types of imitation and referential language: arbitrary, instrumental, body and facial imitation in children with ASD (mean age 48.8 months). Unlike Stone and colleagues' results, relations between body imitation and language were non-significant. Relations between imitation of instrumental actions and referential language were also non-significant. In contrast, relations between imitation of arbitrary actions and language were significant. The contradictory findings between Stone and colleagues' study and Carpenter and

colleagues' study may be attributed to the fact that children in Carpenter and colleagues' were much older and their body imitation task consisted of only two very simple items.

In contrast to the above studies, Rogers, Hepburn, Stackhouse and Wehner (2003) reported no relations between the different types of imitation and language. More specifically, they reported no significant correlations between either postural imitation or imitation of non-meaningful actions on objects and language in children with ASD aged 26-41 months when overall developmental age was partialled out. Nor was a relationship found between the total score of imitation and language. However, the administration procedure used in their motor imitation task was very different from most studies in that physical prompting was provided for children who did not respond to the model. This may have provided strong scaffolding for children's imitation ability, as acknowledged by the authors. Thus, their responses may have not been a true reflection of their imitative ability. In addition, relations were examined partialling out developmental level which was calculated by averaging verbal and nonverbal developmental age.

To summarize, almost all studies of children with ASD have documented an association between motor imitation and language. In most cases relations between imitation skills were only examined with expressive language. Some researchers have suggested stronger relations between body imitation and language compared to imitation of actions on objects (McDuffie et al., 2005; Stone et al., 1997). However, this conclusion was not supported by the findings of Ingersoll and Meyer (2011) and Carpenter et al. (2002). It seems that relations may be influenced by a number of factors such as age, verbal ability and familiarity of the task. Furthermore, it appears that in general relations with language may be stronger when the imitative act serves a social function such as body imitation tasks in which there is no obvious goal or the imitation of arbitrary actions in which the goal is achieved through unconventional means. These types of imitation may depend more on the child's ability to understand others' intentions (Carpenter et al., 1998), form a mental representation of the copied behavior (Stone et al., 1997) and willingness to initiate and sustain interaction with others (Nielsen, 2006). In line with this, Dohmen, Chiat and Roy (2013) argued that poor performance on imitation tasks which mainly serve a social function may be indicative of deficits in the sociocognitive abilities necessary to language acquisition. Children with these difficulties are expected to show specific deficits in language (Chiat, 2001; Chiat & Roy, 2008).

3.4 Motor Imitation in Clinical Groups

This section provides an overview of the performance of children with autism and children with early language delay on motor imitation tasks in comparison to typically developing children. Findings from studies in both groups will provide important information on the range of difficulties that may be observed across both groups. Results will be discussed and implications for the potential of using a motor imitation assessment with clinically referred preschool children with concerns about language will be considered.

3.4.1 Motor imitation in children with ASD – deficit or delay?

Difficulties in imitating other people's movements have been reported in autism in several studies across the past 40 years. The earliest suggestion of a relationship was put forward by Ritvo and Provence (1953). They noted that a mother described her child's inability to imitate her and make pat-a-cake just by watching her. Since that time, different studies have examined imitation abilities in children with autism. Researchers have suggested that imitation abilities differentiate children with autism from children with other developmental disorders. For example, in a study by Stone, Lemanek, Fishel, Fernandez, and Altemeier (1990), ninety one children, aged 3-6 years, in five different groups (autism, mental retardation, hearing impairment, language impairment and typically developing children) were compared on a motor imitation task that included meaningful actions with objects and body movements. Results showed that children with autism performed significantly below all other groups of children. It should be noted that children with ASD had significantly worse verbal skills than all other groups, so we cannot rule out the possibility that the group differences were due to language ability. However, similar findings were reported in a later study in which children with ASD were matched to children with developmental delays (DD) on their chronological, mental and verbal age (Stone, Ousley and Littleford, 1997). Findings revealed that children with autism (aged 26-36 months) performed significantly below DD matched controls on a task that included imitation of actions on objects and body imitation. Unexpectedly, though, this study found that the imitation of children with autism was not significantly different from typically developing (TD) controls matched on mental age. The authors suggested that this might be

due to the TD children being very young (mean age 18.1 months) and possibly less exposed to imitation learning opportunities.

The majority of early studies comparing the imitation abilities of children with autism to typically developing children or other clinical groups have concluded that children with autism show important imitation deficits. In their review of studies, Rogers and Pennington (1991) proposed that imitation represents a primary deficit in autism, leading to a cascade of effects on multiple cognitive and social domains. This view however, has changed in the past 20 years, with a growing body of literature suggesting that evidence points to delayed development of imitation rather than an absolute deficit (see Williams & Whiten, & Singh, 2004, for a review). This conclusion was based on research that found an improvement in imitation abilities with age (Stone et al., 1997; Young et al., 2011) and no group differences in imitation of actions on objects between children with autism aged 5-18 years relative to age matched controls with mental handicap (Charman & Baron-Cohen, 1994). On the other hand, impaired imitation performance on the same tasks was reported in 20-month old infants with autism (Charman et al., 1997). However, it should not be overlooked that the two studies (Charman & Baron-Cohen, 1994; Charman et al., 1997) differed in the matching procedure used between children with autism and the control group: while in the 1994 study children with autism had a higher non-verbal mental age (NVMA) compared to controls, there was no difference in NVMA between the younger children with autism and the DD group in the 1997 study.

In line with the finding that different types of imitation show different predictive relations to language outcome, it has been suggested that the imitation performance of children with autism varies depending on the type of imitation task with some types of imitation showing more impaired performance than others. In general, imitation of actions on objects was reported to be less affected in comparison to imitation of actions without objects (Ingersoll & Meyer, 2011; Stone et al., 1997; Young et al., 2011). In addition, better performance was found on imitation of meaningful tasks in comparison to non-meaningful tasks (Carpenter, Pennington & Rogers, 2002; Stone et al., 1997). Conversely, Rogers and colleagues (2003) reported that children with ASD were more impaired than controls on imitation of actions on objects but not on imitation of body movements. The authors also reported that children with autism did not differ from control groups in their frequency of responding to the model. However, these findings may be due to the fact that,

as reported in the previous section they differed in their administrative procedure by providing physical prompting for children.

Interestingly, the reported dissociations in imitation performance (Ingersoll & Meyer, 2011; Stone et al., 1997; Young et al., 2011) are not unique to children with ASD. A similar pattern has been observed in typically developing children. It was reported that up to two years of age, typically developing children are more likely to imitate actions on objects than gestures (Abravanel, Levan-Goldschmidt, & Stevenson, 1976; Stone et al., 1997). Such results may provide further support for the suggestion that imitative performance in autism reflects a general delay, with the items that are most developmentally difficult showing impaired imitation.

Recently, in a prospective study, Young and colleagues (2011) have suggested that imitation abilities are significantly linked to language abilities and social behavior in children with autism, children with developmental delay and typically developing children and that this relationship follows a similar pattern in all groups. In this study, the development of imitation was examined in a group of typically developing infants and infants at familial risk of ASD first seen between the ages of 12-24 months and followed up at 3 years of age. The imitation task that was used included imitations of actions both with and without objects and oral facial imitations. Increase in imitation ability over time was significantly related to expressive language growth in all groups of children. The authors concluded that their findings cast further doubt on the notion that imitation deficits are specific to autism.

To conclude, there is replicated evidence of poorer performance of children with ASD on imitation tasks in comparison to controls though it has been suggested that this resembles a delay rather than a deficit in the ability to imitate.

However, it is not clear from the results of the above reviewed studies if the poor performance of children with autism on imitation tasks was due to non-compliance or failing to imitate the actions correctly. This was only analyzed in Roger and colleagues' study in which they suggested that children with autism did not differ from control groups in the frequency of responding to the model.

3.4.2 Is Poor motor imitation specific to ASD?

The previous sections have provided considerable evidence of the association between motor imitation and language in both typically developing children and children with ASD. Furthermore, it has been suggested that poor performance on motor imitation tasks may not be specific to children with ASD. Given these findings, it is reasonable to question whether children with language impairment would show differences in their performance on motor imitation tasks in comparison to typically developing children. In fact, there is replicated evidence of impaired performance of children with SLI on motor imitation of body movements (Hill, 1998; Marton, 2009; Vukovic, Vukovic, & Stojanovik, 2010). However, these findings, though interesting, were in studies of much older children than the targeted age range in the current study. Based on the argument that motor imitation may serve different functions during development (Nielsen, 2006) and reflect different underlying processes at different ages, it is more relevant for the purpose of this study to review the evidence from studies of motor imitation with preschool children.

Thal and Bates (1988) examined the ability of 9 late talkers between the ages of 18 and 32 months to imitate meaningful actions on objects and gestures compared to language matched and age matched controls. Late talkers' expressive vocabulary fell in the lowest 10% for their age and their fine and gross motor ability was within the normal range for their age. Results showed that late talkers performed like language matched controls on the imitation task and below age matched controls. Furthermore, children in the three groups did better on the imitation of actions with objects in comparison to gestural imitation.

In a later study Thal, Tobias, and Morrison (1991) followed up a group of late talkers (10 children) one year after the initial assessment. At time 1 the same imitation task used in the previous study was administered (gestural imitation and imitation of meaningful actions on appropriate objects) and the same selection criteria for late talkers were employed. Language assessment at follow-up showed that some of the late talkers appeared to have caught up while a group of them were still delayed in terms of their productive vocabulary. The authors examined whether or not the two groups differed in their performance on the initial assessment measures. Results showed that the two groups did not differ in terms of their productive vocabulary. On the other hand, a significant

difference was found on the imitation task with children who appeared to have caught up performing significantly better than the group who remained delayed.

In contrast to the above findings, Stone and colleagues (1990) found no significant difference in the performance of language impaired children on imitation tasks (mean age 4;5 years) compared to age matched typically developing children. This may have resulted from using a task that was developed for 18-40 months old children (DeMyer et al., 1972) with the much older group of children with language impairment. Most of the items in this measure were imitation of meaningful actions on objects as well as simple or familiar gestures and both groups of language impaired children and typically developing children scored highly on this measure.

Interestingly, in a recent study Dohmen, Chiat and Roy (2013) examined not only whether preschoolers with language delay differed in their performance on imitation tasks compared to age matched controls but also whether their performance was affected by the type of imitation task. Forty-five children with specific language delay ranging in age between 2;0 and 3;5 years were compared to age matched controls on an imitation battery. Participants did not differ in their motor ability based on their performance on a standardized test. The imitation battery included facial imitation, imitation of postures, imitation of gestures, imitation of meaningful (familiar and novel) actions on objects, and imitation of pretend acts on substitute objects. Their task of imitation of pretend acts may be classified as imitation of non-meaningful actions on objects since we cannot be certain that children understand pretense when they are imitating. Results showed that both groups performed near ceiling on the imitation of meaningful actions on objects, whereas significant differences emerged between children with and without language delay on the imitation of postures (for all age ranges), gestures (for the two youngest age groups: 2;0-2;11) and on the imitation of non-meaningful actions (the two youngest age groups). These researchers further investigated whether language delayed children's poor performance was due to not responding to the model or inability to correctly imitate the model. They reported that the poor performance of children with language delay was largely due to not responding and argued that poor performance might be a result of children's inability to understand the examiner's intentions and engage with her.

Taken together, results from the above reviewed studies suggest that like children with autism, young children with language delay may show impaired performance on imitation

tasks relative to controls especially on body imitation. Furthermore, their poor performance cannot be attributed to their motor ability.

3.5 Summary

Studies reviewed in this chapter have demonstrated that imitation plays an important role in the development of language and communication skills. Relations between language abilities and motor imitation in typically developing children and children with autism have generally been found when a measure of imitation of non-meaningful actions on objects or imitation of body movements was used. Significant relations with expressive vocabulary were reported in most studies. Furthermore, researchers have established the poor performance of children with autism on motor imitation tasks in comparison to controls. However, it was suggested that the various types of motor imitation are differently affected in children with autism. In line with the findings of relations between motor imitation and language, it was proposed that children with autism show more impaired performance on the imitation of body movements compared to the imitation of actions on objects. However, these observations are not unique to autism; typically developing children showed the same developmental pattern. In addition, children with language delay showed poor performance on imitation of body movements relative to typically developing children. Collectively, these findings suggest that including a motor imitation test when assessing children with concerns about their language will potentially contribute to the diagnosis and prognosis of children and understanding the nature of their difficulties.

CHAPTER 4

DEVELOPMENT AND ADAPTATION OF ASSESSMENTS

4.1 Introduction

A range of sociocognitive and language measures were chosen to address the aims set out in chapter 1. The battery consisted of six measures assessing sociocognitive and language skills using direct and indirect methods, some existing and some newly developed or adapted for this project. Sociocognitive measures were the Early Sociocognitive Battery (ESB; Chiat & Roy, 2006b), together with a new Motor Imitation test and Sociocognitive Questionnaire; language measures included the Sentence Repetition test (Wallan, Chiat, & Roy, 2011), a new Arabic research adaptation of the Language Use Inventory (O'Neill, 2009), and a preschool adapted version of the Arabic Picture Vocabulary Test (Shaalan, 2010).

This chapter focuses on the measures that have been adapted or developed for this research. The following sections describe the adaptation and development stage, followed by a description of the pilot studies. Participants, test procedures and scoring systems are described in detail in the main study.

4.2 Parental Questionnaires

4.2.1 Language Use Inventory (LUI)

4.2.1.1 Rationale

As reported in chapter 1, the Language Use Inventory was chosen to be among the measures of the battery for its predictive value of language outcome. In Arabic there is no published work on the developmental stages of language use or pragmatic acquisition of Arab children. Adapting the Language Use Inventory would provide preliminary information on how Saudi children develop the ability to use language effectively in social interactions, and lay foundations for future evaluation of its predictive value for their language outcome.

4.2.1.2 Process of test adaptation

Adaptation went through the following steps:

- A forward-translation design was used. In the first step the questionnaire was translated independently by two translators in order to get different perspectives and avoid preferences for certain words or expressions. The first translator was the researcher and the second translator was a speech language pathologist. Both translators spoke Najdi Saudi as their native language, were fluent in English, and had experience in working with young Saudi children and counseling mothers.

- The test preserved the English version as far as possible. Test instructions were translated to Modern Standard Arabic, whereas the examples for children's words and utterances were given in Saudi Najdi dialect as it is the variety that Saudi Najdi children acquire as their mother tongue. The Modern Standard Arabic is acquired later usually through literacy and is spoken in formal or educational settings.

- The two translators reviewed the two versions with two further translators who were bilingual speech language pathologists with an interest and expertise in child language, one of whom was working on a PhD in child language. The purpose was to check further the validity of judgments about the equivalence of the English and Arabic versions. The test was reviewed over two sessions each lasting approximately 3 hours. Group members discussed each item to arrive at the final most appropriate equivalent translation taking into account cultural and linguistic factors.

- Few changes were made to the examples used in the original LUI and these mainly involved food, toys, or transportation items which required substitution or addition of items that are more culturally appropriate in Saudi.

- The researcher reviewed the translated version and edited it so that the language in the test was smoothed out.

- A bilingual Saudi linguist with a PhD in Speech Language Pathology was consulted on the translation of items in part 3-N of the test "how your child builds longer sentences and stories" which assesses the use of connectives, time indicators, modals, and mental state terms. In translating this part it was found that

one word in Arabic would sometimes be the translation for two or three different English words, e.g. “later”, “then”, and “next” all translate as the same word in Arabic; the words, “might” and “could” are likewise encompassed by one word in Arabic. In order to maintain the same number of items in this part, three words in the same semantic categories (time and possibility) were added “not yet”, “now”, and “impossible”. Equivalence of items in the Arabic and English versions in terms of difficulty based on age of acquisition or frequency was not ensured due to lack of literature on Arabic language acquisition or frequency data.

- The adapted version was first trialed with a small group of mothers. The researcher met with four Saudi mothers. Education level of mothers in this group ranged from post high school diploma to bachelor’s degree. Mothers were asked to fill out the questionnaire. The researcher then asked the mothers about the clarity of the statements and appropriateness of examples, and to elaborate on some questions to check their understanding. It took mothers 20-30 minutes to fill out the questionnaire. They stated that the items were easy to read and that the instructions were clear. Only one comment was made regarding the wording of one of the questions in the child’s health section and changes were made to clarify this question.

4.2.2 The Sociocognitive Questionnaire (SCogQ)

4.2.2.1 Rationale

As discussed in chapter 2, individual differences in nonverbal sociocognitive skills are related to subsequent differences in language development. Some of these skills might not be observable or might be difficult to elicit in direct assessment. Furthermore, assessing skills using direct assessment methods and parent reports provide an opportunity to understand parents’ perception of their child’s abilities. In some cases, a clear difference in findings between the two assessment methods may indicate effect of contextual experiences on children’s performance or signal the need for enhancing parents’ observations of their child’s communication abilities (Suen, Logan, Neisworth, & Bagnato, 1995). In Arabic, the only available measure that assesses aspects of socialization is an adaptation of the Vineland Adaptive Behavior Scales (Sparrow, Balla,

& Cicchetti, 1984). However, because it is very broad and covers a wide age range, relatively little information is provided on the socialization domain for children aged 2;0-3;5 years. Therefore, the development of a parent report that specifically targets children's social engagement was considered very valuable especially if used alongside a battery of direct assessments testing similar behaviors.

4.2.2.2 Development of the questionnaire

Items were derived from review of the literature and existing parents' questionnaires such as the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001), Communication and Symbolic Behavior Scales Developmental Profile (CSBS-DP; Wetherby & Prizant, 2002), and Ages and Stages Questionnaire: Social Emotional (ASQ-SE; Squires, Bricker, & Twombly, 2002). Skills that were reported to relate to later language abilities were chosen. Most of these skills assess the same parameters measured in the direct sociocognitive assessments in this study (ESB and motor imitation). The questionnaire included 18 items which cover the domains of social interchange, joint attention, empathy, pretend play and imitation. Appendix A presents the items included in the questionnaire with examples of tests that include similar questions, a brief explanation of available evidence for each item and rationale for its selection.

The first version of the questionnaire included the items in English and Arabic and was sent to professionals for their feedback on translation and wording of items. The second version, including items in Arabic only, was then sent to 10 parents to get their feedback on wording of items, understandability and length of questionnaire. Parents reported that the questionnaire took around 5 minutes to complete and rated the items as clear.

4.3 Direct Measures

4.3.1 Motor Imitation (MI)

4.3.1.1 Rationale

Based on findings from chapter 3, this study aimed to develop a measure for testing motor imitation ability. Tasks reported to be associated with language abilities in most studies were chosen (see sections 3.3.1 and 3.3.2) taking into consideration the normal developmental progression of imitation abilities in typically developing children (see section 3.4.1). Accordingly, the battery consisted of postures and gestures. A number of

imitation measures have been used in previous experimental studies, but they included both actions with objects and actions without objects (Rogers, Hepburn, Stackhouse, & Wehner, 2003; Stone, Ousley, & Littleford, 1997), and some were designed for older children (Ayers, 1980). Therefore, a new motor imitation test was developed for this project.

4.3.1.2 Adaptation and development

Overall procedure was based on previous imitation studies (Beadle-Brown & Whiten, 2004; Stone, Lemanek, Fishel, Fernandez, & Altemeier, 1990). Some of the task items were derived from Beadle-Brown and Whiten (2004) and Rogers and colleagues (2003). Postures and gestures included were a mixture of easy and difficult actions. The gestures in general were more difficult than the postures and mostly required the use of the two hands.

Postures: non-meaningful gestures that do not have semantic associations or conventional communicative meaning:

- Touching back of head
- Interlinking fingers
- Patting elbow with one hand
- Bending index finger
- Wiggling a thumb
- Opening one fist

Gestures: meaningful gestures that carry meaning in their form to symbolize a referent (Capone & McGregor, 2004):

- Pouring
- Flying a plane
- Stirring
- Turning the steering wheel
- Throwing a ball
- Pulling a rope

The gestures chosen for the task mostly represent the actions associated with objects. For example, for the gestures of pouring and throwing a ball, the hands' movements symbolize the action typically performed with the object. On the other hand, the gesture of flying a

plane represents the object itself (plane) in an iconic manner. The pilot study provided the opportunity to check if the chosen tasks were within the physical capability of the target group.

Scoring of children's performance was designed to allow coding of the accuracy of imitation, as it has been reported that imitation accuracy increases with age in typically developing children (Piaget, 1962). Furthermore, children with autism showed reduced imitative precision in comparison to carefully matched typically developing controls (Vivanti, Nadig, Ozonoff, & Rogers, 2008).

4.3.2 Arabic Preschool Receptive Vocabulary Test (APS-RVT)

4.3.2.1 Rationale

As discussed in chapter 1, including a direct assessment of language comprehension is important when assessing children with early language delay. In Saudi Arabia, there are no published tests for receptive language. With regard to available tests in Arabic, Abu Allam and Hadi (1990) published a receptive vocabulary test in Kuwait, but this test was standardized on Kuwaiti children aged 4-16 years. In (2010), Shaalan developed the Arabic Picture Vocabulary Test (APVT) in Qatar as part of his PhD project. The test was administered to 107 children aged 4;6-9;4 years. A second version of the test (see below), was administered by another researcher to children aged 2;0-4;0 years (M. Khater, personal communication, May, 12, 2013). Since this test was tried out with a group of children of a similar age range to Saudi children in this study, it was chosen for adaptation to the Saudi dialect.

4.3.2.2 Adaptation and development

The APVT consists of 132 words belonging to 20 semantic categories (e.g., verbs, animals, occupations, adjectives). Order of the words in the first version of the test was based on the difficulty ranking which was determined by the familiarity rating of 24 adult speakers of Qatari Arabic. Adults were asked to rate 600 words in terms of their difficulty on a rating scale from 1-5 (1 = rarely heard or used, 5 = very frequent and familiar). Out of these 600 words, 132 were chosen and organized into 11 groups consisting of 12 words per group. In the second version of the test, the items were re-arranged based on the

proportion of correct responses of children aged 8;0-9;4 years (Shaalan, 2010). Test items were in the Qatari dialect with some of the vocabulary items in Classic and Modern Standard Arabic.

The test consists of a record form and 134 page booklet with the first two pages for the practice items and 132 pages for the test items. Each page consists of a set of 4 black and white line drawings that were mostly taken from the BPVS (Dunn, Dunn, Whetton, & Burley, 1997) or from non-copyrighted material (e.g., free clip art). In order to reduce fatigue effects, a ceiling criterion of eight errors in one group was imposed: once reached the test was stopped.

In order to adapt the test to be used with young Saudi children a number of changes were made:

- Since this study targeted very young children (aged 2;0-3;5 years), only the first eight groups of the test were included. Thus, in its first pilot version the test consisted of 96 vocabulary items and two practice items.
- The researcher drew the pictures for all the items. This ensured that stimuli reflected cultural conventions in terms of people's clothing, actions and objects, and were familiar and appropriate for young children. It also ensured consistency in form and style of pictures (see Figure 1).

Figure 1: Examples of new pictures stimuli of the APS-RVT

- Saudi Najdi dialect was used for all the items. Three native speakers of Najdi dialect were consulted on the equivalent Najdi terms to the words originally in Qatari dialect or standard Arabic.
- Items which were culturally inappropriate or used rarely in Saudi were replaced by other items (e.g. wild pig was replaced with zebra, and ibex was replaced with goat).
- The word "judge" was substituted by "engineer" because judges do not wear special clothes representing their profession in Saudi and hence cannot be represented in a picture that children could identify.
- Cultural considerations led to some changes of the foil pictures (e.g. the picture of a woman delivering mail was replaced by a woman giving a gift). In choosing an alternative foil picture care was taken to ensure that the same word in Arabic would be used to describe the substitute picture.
- The word "cat" was listed twice in the original test, once in the Qatari dialect and the second time in Standard Arabic. Since only the Najdi dialect was used, the word "tiger" was used to substitute the word "cat" in Standard Arabic.

4.4 Piloting

Two pilot studies were conducted to check that the battery was informative and manageable with young children, and to address the following:

- Time needed to complete the assessments.
- Response rate of mothers completing the questionnaires.
- Compliance of children on the direct assessments.
- Appropriateness of order of presentation of assessment tools.
- Appropriateness of items in the Motor Imitation test.

The pilot study also provided preliminary data on children's performance on newly developed and adapted tests, and allowed preliminary comparison of children's performance on the different measures. For ethical approval see section 5.1.

4.4.1 First pilot study

In the first pilot study, parents were asked to complete the non-standardized Arabic Research Adaptation of the Language Use Inventory (ARA-LUI) and the Sociocognitive Questionnaire (SCogQ), and the researcher administered the direct assessments in the following order: ESB, Motor Imitation test (MI) and the Arabic Preschool Receptive Vocabulary Test (APS-RVT). No ceiling criterion was used on the APS-RVT and children were tested on all the items in the eight groups in order to determine the appropriateness of the ordering of the items. The Sentence Repetition test (SR) was not administered in the piloting stage since it had been developed and tested with Saudi children (Wallan, Chiat, & Roy, 2011).

4.4.1.1 Pilot participants

The pilot study was carried out with 19 children in a nursery in Riyadh, Saudi Arabia. Parents of children aged 2;0–3;5 were sent an information letter about the project and consent forms. Questionnaires (ARA-LUI, SCogQ) were then sent to those who agreed to participate by signing the consent form. Thirty-six questionnaires were distributed. However, only 23 were returned. At this stage only children whose parents completed the questionnaires were included as one of the aims was to evaluate relations between children's results according to parental reports and direct assessment measures. Out of the 23 completed questionnaires, four children had to be excluded for a number of reasons: two children declined to join the researcher and participate in the tests and two children stopped attending the nursery. Thus, the final number of participants was 19. All children spoke Saudi Arabic as their first language. Table 3 presents the age and gender of children included in the first pilot study.

Table 3: Participants in the first pilot study

Age group	Age in months	Girls	Boys	Total
1 (2;0-2;5)	26	-	1	1
	27	1	-	1
	28	-	2	2
	29	2	-	2
Total age group 1		3	3	6
2 (2;6-2;11)	31	1	-	1
	32	1	2	3
	34	1	-	1
	35	1	1	2
Total age group 2		4	3	7
3 (3;0-3;5)	36	-	1	1
	37	-	1	1
	38	1	-	1
	40	2	-	2
	41	1	-	1
Total age group 3		4	2	6
Total		11	8	19

4.4.1.2 Outcomes of the first pilot study

- Only 63.9% of mothers who agreed to participate in the study completed the questionnaires. To increase response rate it was decided to send written reminder notes to mothers in the main study.
- Children were assessed in 1-2 sessions depending on the child's willingness to continue, and for most, these took about 45 minutes to complete.
- The order of presentation of tests during the pilot study was found to keep children engaged. Starting the test battery with the ESB helped children feel at ease and participate with the researcher in the following tasks.
- Children participated in all the direct assessment tasks and appeared to enjoy them.
- Observation of children's performance in the MI test showed that some of the actions were difficult for the children in the youngest age group, but they were within the capability of some of the children in the oldest age group, indicating that they were appropriate for the target age group and had potential to show differences in performance according to age.
- All age groups were able to follow the instructions of the APS-RVT, although the youngest group appeared to lose interest faster than older children. Given that no

ceiling criterion was used during piloting, it was concluded that the APS-RVT was an appropriate measure for the age range. A number of changes indicated:

  - The item "watermelon" has 3 different synonyms in Saudi dialect. This means children may fail the item because they use a different name. This item was replaced with "orange" as both are common fruits in Saudi, but "orange" does not have a synonym.
  - The word "neck" appeared to be difficult for young children to identify in pictures. Although during testing some of the children pointed to their necks when asked about the word, they couldn't identify the word in the picture. This word was substituted with the word "knee".
  - Changes were also made in the pictures for targets "doctor" and "few" to make them clearer.
  - On some items, children appeared to rely on the morphological form (feminine or masculine) as a cue to the target word. For example, the pictures for the target word "doctor" were a male doctor, a female laboratorian, a male electrician, and a male repairman. Feminine form in the foil pictures may reduce the chance from four to three options. Foil pictures were substituted so that they matched the gender of the target word for the items "doctor", "tired".
  - Children showed a tendency to point to a certain position in the page when they got tired or bored. In order to reduce the possibility of such response biases, special care was taken to ensure that in each group target pictures were equally and randomly distributed among the four positions.
  - Based on the accuracy scores of the Saudi sample on the first pilot study, the test was re-arranged with words ranked in decreasing order of correct responses.

- Looking at the mean scores of the three age groups on ARA-LUI, ESB, MI, and the APS-RVT showed that scores increased with age. In contrast, scores on SCogQ were not affected by age. The results of the first pilot study can be found in Appendix B.
- Examining correlations between measures partialling out age in months showed large significant correlations between the following measures: ARA-LUI and APS-RVT, ESB and APS-RVT, MI and APS-RVT, MI and ARA-LUI. The SCogQ did

not show any significant correlations with any measure. Considering that this is a very small group significant relations may be found with a larger more heterogeneous sample.

- Since the APS-RVT was modified after the first pilot, it was trialled again in a second pilot study.

4.4.2 Second pilot study

After modifications to the APS-RVT, the test was administered a second time to 17 out of the 19 children who participated in the first pilot (one child refused to be tested again and another did not complete the test). Children were tested only on the APS-RVT.

Percent of correct responses on all items during the first and second pilot testing were calculated and compared. Percent of children passing individual items in the last group (i.e. group number 8) ranged from 0-6% on both trials, apart from two items which showed 10% accuracy on the first trial. It was concluded that the last group of items would not be informative with very young children or discriminate between young children's lexical abilities. The last group was therefore eliminated. This had the benefit of reducing test administration time and testing children only on items of suitable difficulty. However, one item from group 8 "few" was retained and was substituted for the word "van" from group 7 which showed the same poor performance as items in group 8.

Since there were a number of changes in stimulus pictures and vocabulary items after the first pilot test, the results of the second pilot test were used to determine the final order of the words in the main study based on number of correct responses on each item (see Appendix C).

CHAPTER 5

METHODOLOGY

5.1 Participants

Ethical approval was obtained from the School of Health Sciences Research Ethics Committee at City University London.

The goal was to recruit 150 children whose first language was Saudi Arabic, ranging in age from 2;0 years to 3;5 years and equally divided between boys and girls and between three age groups 2;0-2;5, 2;6-2;11, 3;0-3;5.

5.1.1 Recruitment procedure

Children were recruited from four nurseries (three public and one private) in Riyadh, Saudi Arabia. These nurseries were located in the north, south, west and central parts of Riyadh. Heads of nurseries were sent invitation letters explaining the project (see Appendix D). Those who were willing to participate were then met personally to further explain the purpose of the project and the test battery in general. The researcher worked with nursery supervisors or registration staff to identify children in the targeted age range and matching the inclusion criteria. Parents of children who met the inclusion criteria were sent the information sheet, parental consent form and a questionnaire requesting demographic information and information about the child's medical health and exposure to languages (see Appendix D). This provided a second source of evidence that participating children fitted the inclusion criteria. The number of parents who refused to participate is not known as the researcher had no access to this information. Only children who agreed to join the researcher and participate in the activities were included. A wrong birth date was provided initially for some children by the mother. To confirm accuracy of children's birth date the researcher reviewed the birth dates of all participating children with nursery staff.

5.1.2 Inclusion criteria

- Saudi Arabic as primary language, with child exposed to Arabic since birth.
- Absence of a diagnosis of developmental disorder, autism, or a neurological disorder of known etiology.
- No evidence of severe visual impairment.
- No evidence of hearing impairment.
- No motor impairment.

Children were not excluded if the parent reported that they were exposed to other languages a significant amount of the day as children from the middle high class in Riyadh are becoming increasingly exposed to English from a very young age. Furthermore, mothers were not always accurate in estimating the amount of time of exposure to different languages. However, if the child showed during the assessments that they were not using Arabic and were using another language spontaneously and dominantly, they were excluded because this clearly affected their performance in the APS-RVT and the Sentence Repetition test.

Children with a diagnosed or suspected speech or language problem were not excluded from the sample for a number of reasons. First, information on diagnosed problems came from parental reports and parents might have included children who showed earlier delays and caught up or children who had speech-only problems. Second, clinical judgment of language delay in Saudi is not based on performance on standardized assessments as developmental norms are not available in Arabic.

Based on parents reports 2.48% of the sample were diagnosed with a speech or language delay and 9.94% were suspected to have speech or language delay. Information on language status based on parental reports was missing for almost 16.77% of the sample. Table 4 shows the number of participants in each age group and their language status based on parental report.

Table 4: Participants language status based on parental reports

	No concern			Suspected			Diagnosed			Missing Information		
Age group	G	B	Total	G	B	Total	G	B	Total	G	B	Total
1 (2;0-2;5)	21	19	40	2	4	6	0	0	0	2	3	5
2 (2;6-2;11)	22	20	42	0	1	1	0	2	2	6	4	10
3 (3;0-3;5)	16	16	32	4	5	9	0	2	2	10	2	12
Total	59	55	114(70.80%)	6	10	16(9.94%)	0	4	4(2.48%)	18	9	27(16.77%)

G: girl; B: boy

5.1.3 Excluded children

Twenty children were excluded for the following reasons: four children were not willing to join the researcher; one child did not respond to any direct measure of assessment, had a missing SCogQ and scored 0 on the ARA-LUI; one was hyperactive and did not cooperate during assessment tasks; two children were older than 3;5 years; six could not be tested because they stopped going to school after parents returned the questionnaire; one child had a missing birth date and mother could not be contacted; five children spoke mainly in English during testing. Thus, just six children were excluded for non-cooperation or floor performance.

5.1.4 Participants age and gender

The final sample included 83 girls and 78 boys, all with Saudi Arabic as their first language. However, a full data set was not available for some participants due to non-return of questionnaires (see Table 5). Children who refused to participate in the MI or SR tasks were included because exclusions would risk losing important information about children's imitation performance (Dohmen, Chiat & Roy, 2013). For further discussion of children's compliance see section 7.1.3.

Table 5: Number of participants in direct measures and parental questionnaires according to gender and age group

		Girls	Boys	Total
Age group 1 (2;0-2;5)	Direct Assessments	25	26	51
	Returned ARA-LUI	23	23	46
	Returned SCogQ	22	23	45
Age group 2 (2;6-2;11)	Direct Assessments	28	27	55
	Returned ARA-LUI	23	23	46
	Returned SCogQ	24	23	47
Age group 3 (3;0-3;5)	Direct Assessments	30	25	55
	Returned ARA-LUI	20	22	42
	Returned SCogQ	20	23	43
	Total	83	78	161

Notes: (i) All participants were included in direct measures except for one child on the APS-RVT from age group 3 due to non-cooperation. (ii) SR was administered to age groups 2 and 3 only. (iii) Most parents either returned the two questionnaires (ARA-LUI, SCogQ) or neither. However, in very few cases one questionnaire was fully completed and the other was not.

5.1.5 Demographics of the participants

Parents were asked to specify their educational level (high school degree, diploma, bachelor, postgraduate degree, other). Based on educational level, most children came from middle class families. Just under half of the fathers had completed their bachelor's degree (39.75%), while the other fathers' education ranged from illiterate (0.62%) to those who held a postgraduate degree (19.25%). With regard to the mothers' education, 10.56% had completed high school, 57.76% had a bachelor's degree, and 19.26% had a postgraduate degree. Only 1.24% and 3.11% had completed only primary school and only intermediate school respectively.

It is not possible to determine how the sample of participants in this study compares with the Saudi population since comparable data are not provided by the Saudi Central Department of Statistics and Information (CDSI). However, the CDSI does provide data on the percentage of Saudi employees by educational level in 2009 (see Appendix E). Table 6 compares the educational level of parents of participants to the educational level of employed Saudis in 2009. Employment status of parents of our sample was as follows: 86% of fathers were employed and 64% of mothers were employed, while the other mothers were either students (17%) or unemployed (14%).

Table 6: Breakdown of educational level for parents of participants compared with employed Saudis in 2009

Education Level	% Parents of participants		% Employees in Saudi 2009	
	Female	Male	Female	Male
Illiterate	0.00	0.62	2.10	4.10
Below high school	4.35	4.97	5.60	35.50
High school	10.56	21.12	10.40	31.80
Diploma	4.97	9.94	17.20	7.90
Bachelor and over	77.02	59.00	64.60	20.70

The educational level of 3.11% of mothers and 4.35% of fathers in our sample is unknown due to non-responding

As shown in Table 6, the educational level of mothers in our sample was broadly similar to that of employed female Saudis, though a higher proportion had a university degree. However, fathers in our sample appeared to have a higher educational level in general. Although our sample may not be truly representative of the Saudi population, characteristics of parents in terms of educational level and employment status are very much in line with parents of young children seen in speech and language clinics. Based on clinical experience, parents who seek speech and language therapy services for their young children in Saudi usually come from middle to high socio-economic backgrounds, though there are no published statistical figures yet from speech and language clinics (see section 7.5.1 for further discussion of representativeness of the sample).

5.2 Procedure

5.2.1 General procedure

Children whose parents agreed to their participation by signing the consent forms were sent the non-standardized research adaptation of the Language Use Inventory (O'Neill, 2009) into Arabic (ARA-LUI) and the Sociocognitive Questionnaire (SCogQ). Parents were sent reminder notes if they did not return the questionnaires after a week. Direct assessments were administered within a week of return of parental questionnaires and in some cases the mother completed the questionnaires while attending the direct assessment session. In order to minimize bias, the researcher was not aware of parents' responses and scoring of questionnaires was done after the completion of direct assessments.

Testing on direct measures was carried out in one or two sessions depending on the child's cooperation and willingness to continue, in a room in the nursery. Completion of all assessments lasted approximately 30-45 minutes. When possible the child's teacher accompanied the child during testing and in a few cases the child's mother. Testing rooms varied in space, level of noise and number of toys in the room. In some cases the only available room was the nursery's resource room which was full of toys and distracters. This made it difficult to test some children as they were more interested in exploring the different toys in the room, and consequently it meant longer testing time. However, once testing started children appeared engaged and attended to the tasks.

The session started with the researcher introducing herself to the child and explaining that they would play some games together. Then the test battery was administered in the same order as the pilot study with the exception that the Sentence Repetition test (Appendix F), which was not administered during piloting, was administered after the Motor Imitation test (MI) in the main study. Thus the order of the battery was as follows: the Early Sociocognitive Battery (ESB; Chiat & Roy, 2006b), the Motor Imitation (MI), Sentence Repetition (only for children between 2;6-3;5 years), and the Arabic Preschool Receptive Vocabulary Test (APS-RVT). For most children, this order was maintained. However, five very shy children were not cooperative with the researcher during the MI and SR tests on first attempt so the researcher administered the less demanding vocabulary test (APS-RVT) first and ended with the MI and SR tests. Children's performance was scored live on record forms (Appendix G) except for the SR test which was audio recorded. Children were rewarded with a sticker at the end of the session.

5.3 Assessment Battery

5.3.1 Parental questionnaires

5.3.1.1 The Arabic Research Adaptation of the LUI (ARA-LUI)

The adapted version closely resembled the original English version. As described in section 4.2.1.2, test instructions are in standard Arabic, the variety usually used in written discourse, while Saudi Arabic (Najdi dialect) is used for the examples. The questionnaire includes a total of 180 questions most of which require yes/ no answers. It consists of 3

major parts: “How the child communicates with gestures”, “How the child communicates with words”, and “the child’s longer sentences”.

In accord with the original test, parents were instructed to complete all the questions in 1-2 days and consult with other family members if needed. The same rule of dealing with missing data in the original test was also used: when the answer for more than two items for scored subscales was missing, the total score for the questionnaire was not calculated and the questionnaire was excluded.

The maximum raw score on the LUI is 161, which is the sum of points in Parts 2 and 3. The first part of the test which focuses on communication with gestures is not included in the total score. For children with very low verbal ability, information can be obtained from the first part on how they communicate nonverbally.

5.3.1.2 Sociocognitive Questionnaire (SCogQ)

This newly designed questionnaire elicits parents’ observations of their child’s communicative and social skills. It includes 18 items, covering the domains of social interchange, joint attention, pretend play, imitation and empathy (see Appendix G). Parents were instructed to mark the answer that best describes their child’s behavior based on their daily observations.

Items were scored on a 3 point scale (2, 1, 0) to designate whether the child often exhibits the behavior (2 points), sometimes (1 point) or never (0 points). Three items (no.13, 14, and 17) were negative and were reverse-scored, for example, “prefers to play alone”, so that 2 points are awarded when the parent indicates that the child never exhibits the behavior, 1 point for sometimes, and 0 for often. A total score was generated by summing the scores for the 18 items. The maximum total score for the questionnaire = 36.

5.3.2 Direct measures

5.3.2.1 Early Sociocognitive Battery (ESB)

The Early Sociocognitive Battery developed by Chiat and Roy (2006b) is a set of nonverbal tasks which have been found to relate to language and communication

development (Chiat & Roy, 2013). It includes three tasks which assess the child's reaction to the researcher's expressions of feelings; their ability to share attention by monitoring the researcher's gaze; and their understanding of gestures and objects used as symbols.

These assessments were administered in a master's project (Alkadhi, 2010) with 35 middle to upper class Saudi children. The test script was translated into Saudi Arabic and the test materials described in the protocol (http://www.city.ac.uk/lcs/research/veps/assess.html) were used as they were all considered culturally appropriate, apart from one item in the practice condition of the symbolic comprehension task (rattle), which was changed to a car because the word rattle in Saudi varies between regional dialects. Results showed that young Saudi children participated well in these tasks and appeared to enjoy them. Furthermore, comparison of results between Saudi children in Alkadhi (2010) and children in the UK in Chiat and Roy (2006a) showed very similar scores despite the difference between the sample of children and testing situations in the two studies. In terms of British cut-offs, the percentage of Saudi children falling in the normal range was very similar on all three ESB tasks.

Full ESB protocols are available on: http://www.city.ac.uk/lcs/research/veps/assess.html

For the reader's convenience the assessment procedures and scoring criteria are briefly described below:

1. Social responsiveness:

This non-verbal task is based on a procedure developed by Sigman, Kasari, Kwon and Yirmiya (1992).

In this assessment, the examiner expresses six different feelings (hurt, surprise, anger, fear, distraction, and achievement) in a sequence of scenarios. Expressions are exaggerated and last up to 5 seconds. The child's response to the examiner's expression of feeling is measured by the child's gaze to the examiner's face.

The child scores 2 points for looking at the examiner's face for at least 2 seconds during expression of feeling and 1 point for the child's fleeting look. No points are given if the child does not look at the examiner's face. Maximum total score for this task = 12.

2. Joint attention:

In this assessment, the child's gaze switch from an object to the examiner (or vice versa) and the child's gaze or point following are measured. To measure gaze switch, the

examiner monitors the child's eye gaze at two points. First the examiner shakes a plastic egg containing a miniature object. Then the egg is opened revealing the miniature object, and the child is given time to play with the miniature toy. The child is awarded one point for alternating gaze either between the egg and the examiner's face before the egg is opened or between the small object and the adult after the egg has been opened. No points are given if the child does not switch gaze at either of the above opportunities. Next, the adult looks towards an object matching the miniature in the egg, and the child's gaze monitoring is measured. Two points are awarded if the child follows the examiner's direction of gaze. If not, the examiner points at the object and one point is awarded if the child follows the examiner's finger point. No points are scored if the child fails to respond to either gaze or finger-point. Maximum total score for the joint attention task = 18.

3. Symbolic comprehension:

This task is adopted from a procedure developed by Tomasello, Striano, and Rochat (1999).

In this essentially non-verbal task the child's understanding of different levels of symbolic comprehension is tested.

The task takes the form of a game of finding an object and rolling it down a chute. The task starts first with a practice condition where the examiner asks the child to roll a named object down the chute. The aim of the practice condition is to make sure that children understand the examiner's communicative intentions in the context of this game. However, the child's correct identification of items is not required for the continuation with the assessment conditions.

The three conditions of symbolic representation in this task are: gesture, miniature object, and substitute object which are used as symbols for a target object.

- In the gesture condition the examiner gestures an action appropriate for the requested object.
- In the miniature condition the examiner holds a replica miniature item to ask for the full sized item.
- In the substitute condition the examiner acts symbolically on an object transforming it to another object, for example use a banana as if it was a telephone, then use the real object to ask for the objects that were acted on symbolically.

The child scores one point for the correct selection of the target object from a set of six in each of the three conditions. Maximum score for the total symbolic comprehension task = 18.

5.3.2.2 Motor Imitation (MI)

This newly designed test assesses the child's ability to imitate non-meaningful postures and gestures modeled by the researcher (see section 4.3.1.2).

The child was seated facing the researcher. The researcher explained to the child "we are going to play a game. I will do an action and then you do whatever I did". Then the researcher modeled two practice actions one at a time (touching earlobe, grabbing nose) and after presenting the action the child was verbally prompted "do what I did" if he/she did not imitate spontaneously. If the child did not respond to the first presentation of actions during the trial period, the researcher modeled the actions a second time and physically helped the child carry out the action to ensure that the child understood the nature of the task. The practice actions were then presented one more time and the child's response during this time was required for the continuation of the test.

When the child imitated or attempted to imitate the researcher during the practice period, the researcher presented the 6 test postures one at a time followed by the 6 gestures. Each posture was held for approximately 3s. After the researcher modeled the target action, the child was instructed "your turn".

If the child became distracted at any time the researcher attempted to redirect the child's attention before demonstrating the next action by calling the child's name and saying "watch what I am doing" or "look at me". If on any action the child did not imitate the first demonstration, the researcher modeled the action a second time and verbally prompted the child "do what I did". If the child did not respond after the second presentation then a "no response" was coded and the next action was demonstrated. Usually the child's failure to respond to the first model was due to distraction, thus a second model was allowed throughout the test when needed.

The child was rewarded for any attempt with clapping and /or praising the child "well done" or "good job!", or was given a piece of a simple puzzle.

Response accuracy was scored live on a 3 point scale. Two points were awarded for an accurate response, one point for partial success and no points for failure or no response. Partial success meant that the child attempted to imitate the adult's action but made an error in the location, form, or movement of the action. Children were free to use their right or left hand since it has been reported that young children tend to use their dominant hand while imitating an act (Gleissner, Bekkering, & Meltzoff, 2000).

Practice items did not count towards total score. Maximum score for the MI task is 24 which is the sum of the total score for posture imitation (12 points) and gesture imitation (12 points).

5.3.2.3 Sentence Repetition (SR)

The Sentence Repetition test was developed by Wallan, Chiat and Roy (2011) and administered to 140 children aged 2;6-5;11 years in Riyadh, Saudi Arabia. The test consists of 14 sentences in Najdi dialect ranging in length from 5-7 words. Sentences were created specifically for Arabic and consist of a variety of basic sentence structures with key grammatical markers (see Appendix F).

Children were given the following instructions "I will say a sentence, I want you to listen carefully until I finish the sentence and then repeat the sentence exactly like I said it. Are you ready?" Then two practice sentences were presented to familiarize participants with the test. If the participant gave an incorrect response or did not respond, the researcher repeated the instructions and modeled the correct response then re-administered the practice sentences. The 14 target sentences were presented if the child responded correctly to the first practice sentence or after the administration of the second practice sentence, irrespective of the child's response. Sentences were arranged in order of increasing length in grammatical morphemes, and order of presentation was fixed. If the child did not respond to a target sentence on the first presentation, the sentence was presented a second time to allow one further opportunity to repeat.

Practice sentences did not count towards the participant's score. Any misarticulations or changes of word order from Verb-Subject to Subject-Verb or vice versa were not considered errors since the order of these elements is free in Arabic. For each participant the following three scores were obtained: lexical morpheme score (maximum score = 56),

grammatical morpheme score (maximum score = 117), and total sentence accuracy (maximum score = 42). For full details of the Sentence Repetition test items and scoring (see Appendix F).

5.3.2.4 Arabic Preschool-Receptive Vocabulary Test (APS-RVT):

This test was adapted from the APVT (Shaalan, 2010) as described in the previous chapter. The new test included a picture booklet and a record form. The record form contained 84 vocabulary items all in Saudi Arabic. Items were equally divided into 7 groups each consisting of 12 items. Item order was determined by the second pilot study based on number of correct responses each item received. The pictures booklet consisted of 86 pages, each page showing four black and white drawings. The first 2 pages were for the practice items and 84 pages for the test items. Target pictures in each group were equally and randomly distributed among the four positions to reduce possibility of response bias.

The tester presented the pictures and asked children to point to the picture corresponding to the orally presented stimulus word using the following instructions in Saudi Arabic "now I will show you some pictures…where is… show me.." before the presentation of each item. Practice items were presented first to ensure that children were able to follow the examiner's instructions and point to one of the four pictures on the page. Correct identification of practice items was not required. Practice items were not scored. Once the child made 8 errors in a group that group was completed but testing did not proceed to the next group.

Children's responses were recorded on the record form. The following coding system was used: correct responses (✓), wrong responses (number of the picture that child pointed to), and no response (NR). One point was awarded for a correct response and 0 for incorrect responses or no response. The total raw score was computed by subtracting the number of errors from the number of the last tested item. For example if the last tested item was item number 36, and the total number of errors was 11, the raw score would be 25. The maximum possible score is 84.

CHAPTER 6

RESULTS

This study aimed to develop a battery of early sociocognitive and language measures reported to be good predictors of children's language and communication skills to use with young Saudi children. In addition, it aimed to investigate relations between the sociocognitive and language measures used to assess early skills in young Saudi children. Examining these relations will provide additional insight into relations between sociocognitive skills and language and contribute to our understanding of early language development. Six assessments were included: the Arabic research adaptation of the Language Use Inventory (ARA-LUI), the sociocognitive questionnaire (SCogQ), the Early Sociocognitive Battery (ESB; Chiat & Roy, 2006b), the Motor Imitation test (MI), the Sentence Repetition test (SR; Wallan, Chiat, & Roy, 2011), and the Arabic preschool receptive vocabulary test (APS-RVT). The results chapter starts with examining the reliability and validity of the assessments. In the second section, the effect of gender is examined to determine if gender should be taken into account in analyses. In the third section descriptive statistics for all the tests in the three age groups are presented as well as inferential statistics examining the effect of age. This is important as one of the ways of assessing the potential use of the measures in differentiating between typically developing children and children with language or communication problems is to investigate whether these measures reveal differences in performance between different age groups. In addition, results are considered in relation to previous research findings and key issues identified for further discussion. The fourth section examines associations between the different sociocognitive measures and language measures using correlational analyses. Then, the unique contribution of sociocognitive skills to different aspects of language is investigated using regression analyses. The fifth section examines the pattern of performance of low performing children and identifies different profiles. Finally, in the last section, a comparison is made between parental report of concern and children's performance on the battery of assessments.

6.1 Reliability and Validity

Reliability and validity are psychometric aspects of a test that are important to address as part of the development of a new test before the interpretation of test results. Reliability and validity of tests adapted in this study were examined in previous studies (Chiat & Roy, 2006; O'Neill, 2009; Shaalan, 2010). For example, Chiat and Roy (2006) reported high levels of inter-rater reliability on the three tasks of the ESB, with intraclass correlations of .9-.96. In addition, intercorrelations between the three measures of the ESB (social responsiveness, joint attention, and symbolic comprehension) ranged from r = .3 to .6, when age in months was controlled. Likewise, O' Neill (2009) and Shaalan (2010) results support the reliability and validity of the Language Use Inventory (LUI) and the Arabic Picture Vocabulary Test (APVT) respectively. Good levels of internal consistency, test-retest reliability and other forms of reliability and validity were reported in both studies. However, since some of those measures were administered in a different language culture and some measures were largely modified in the adaptation process, it is important to establish their reliability and validity with young Saudi children (van de Vijver & Poortinga, 2005). In addition, one must take into consideration in using these measures not only the linguistic equivalence and cultural appropriateness of the test items, but also how participants in such a very different culture respond to the assessments especially with regard to parents filling in questionnaires. In contrast, reliability of the Sentence Repetition task was established recently in a previous study with Saudi children (Wallan, Chiat & Roy, 2011). Near perfect levels of reliability were reported for the three scores of Sentence Repetition (A. Wallan, personal communication, October, 25, 2014) on both inter-rater reliability (α =.99 for lexical morpheme score, grammatical morpheme score and total sentence accuracy score) and test-retest reliability (lexical morpheme score =.98, grammatical morpheme score = .99, total sentence accuracy score = .96), thus it was not assessed again in this study.

6.1.1 Reliability

Reliability refers to the extent to which a test performs in a consistent way. This implies that changes in scores produced by the test should only reflect a change in the variable being measured by the instrument (DeVellis, 2012). Reliability can be measured by different techniques such as internal consistency, inter-rater and test-retest reliability. *Internal consistency* measures the degree to which items in a test or subtest hang together

(DeVellis, 2012). Based on the conventions reported in Field (2013), values between .7 and .8 represent good levels of reliability. Internal consistency was measured for the two questionnaires (ARA-LUI and SCogQ).

Other ways of measuring reliability include *inter-rater reliability* which is often necessary when data are collected through ratings or scorings of examiners. Inter-rater reliability measures the correlation found between two different independent raters or examiners (Rust & Golombok, 2009). In the case of ratings of behaviors in a direct testing situation such as in imitation studies, reliability is usually checked through scoring videotaped sessions. However, that was not possible in this project due to the general cultural unacceptability of video recording children for research purposes in Saudi. Accordingly, in this study, inter-rater reliability was established with a second examiner attending the live session. The second examiner had previous experience with the administration of the ESB. She was a speech language pathologist who speaks Najdi Saudi as her native language and who was also working on a PhD in child language at the time the tasks were administered. Prior to the sessions to be scored, the researcher explained the scoring criteria of the Motor Imitation task (MI) and the Arabic Preschool-Receptive Vocabulary Test (APS-RVT) using scoring sheets which included a brief description of the scoring criteria. In addition, examples of the most common errors children perform and how to score those errors were demonstrated to the second examiner. The second examiner attended 10% of the assessment sessions, was blind to the scores of the first examiner and independently scored these children.

Finally, *test-retest* is another form of reliability that is commonly used. Test-retest examines the correlation between the scores of the same test when administered twice to the same subjects within a time interval that differs depending on the type of the test (Rust & Golombok, 2009). Previous assessments of language reported test-retest being measured with an average of 4 weeks interval between tests (Dunn & Dunn, 2007; O'Neill, 2007). An important consideration is that the span is not too long as this might lead to changes in scores as a result of maturation or spontaneous recovery. In this study test-retest varied from a week to 3-4 weeks depending on the measure and availability of children, though in the case of the APS-RVT, the test was re-administered within a week to ten days. Memory effects were not a concern with the re-administration of the APS-RVT within a short period as children were not provided with the correct response and were reinforced always for responding whether or not their response was correct. Test-

retest reliability was measured for 10% of the sample for the parental questionnaires and 14% of the sample for the MI and APS-RVT.

Inter-rater and/or test-retest reliability was measured using a two-way mixed, absolute agreement intraclass correlation coefficient (ICC) with 95% confidence interval. ICC is the preferred statistics used for ordinal, interval and ratio variables (Hallgren, 2012) and a two-way mixed model was chosen since the participants were randomly selected but the second coder was not. In addition, an absolute agreement reliability type was chosen since we were interested in the absolute agreement value between the scores rather than degree of similarity in rank order between scores for which consistency is usually used. ICC values from .60 to .74 indicate good levels of agreement and values from .75 to 1.0 indicate excellent levels of agreement (see Cicchetti, 1994). Reliability results are presented for each test separately in the following sections.

6.1.1.1 Parental questionnaires

Parental questionnaires included the Arabic Research Adaptation of the LUI (ARA-LUI) and the Sociocognitive Questionnaire (SCogQ). Internal consistency was measured for the two questionnaires. Cronbach's alpha values for the two scored parts of the ARA-LUI were .93 and .98 for part 2 and part 3 respectively, and the Cronbach's alpha for the SCogQ was .71, indicating good levels of internal reliability for the two questionnaires.

In addition test-retest reliability was measured. Thirty parents agreed to complete the questionnaires twice within 3-4 weeks, however, only 14 ARA-LUI questionnaires and 15 sociocognitive questionnaires (10% of the sample) were returned twice with complete data. ICC were $\alpha = .97$ for the ARA-LUI and $\alpha = .77$ for the SCogQ indicating that both questionnaires were stable over time.

6.1.1.2 Direct measures

Early Sociocognitive Battery (ESB)

In order to ensure consistency of scoring, 16 children (10% of the sample) from the three age groups were independently scored by a second examiner. Reliability for the total ESB score was calculated as well as for the three measures of the ESB. ICC for the total ESB was .99 which indicates high level of agreement. Similarly, excellent values were obtained for the individual tasks of the ESB, social responsiveness (.87), joint attention (.97) and symbolic comprehension (1.00).

Motor Imitation (MI)

Inter-rater reliability was estimated based on the scores of 16 children (10% of the sample). ICC values indicated a strong level of agreement for the total MI task (α =.96), the posture task (α =.97) and gesture task (α =.92).

In addition, test-retest reliability was measured by retesting 22 children (14% of the sample) 7-10 days after they were assessed the first time. Results showed high levels of agreement for the total MI task (α =.95) as well as for individual tasks: postures (α =.90), gestures (α =.88).

Arabic Preschool Receptive Vocabulary Test (APS-RVT)

Inter-rater reliability was measured based on the scores of 16 children (10% of the sample). ICC value indicated perfect agreement (α =1.0).

Test retest reliability was also measured by administering the test twice to 22 children (14% of the sample), with 7-10 days between the two administrations. An ICC value of .91 was obtained indicating that the test was stable across administrations.

To summarize, it can be concluded from the above reported results that all the measures used in this study are reliable, with excellent levels of agreement achieved on all measures.

6.1.2 Validity

Validity refers to the degree to which a test is measuring what it is intended to measure (Rust & Golombok, 2009). Validity of a test can be established by different forms of validity. Most commonly it includes face validity, content validity, concurrent validity, and construct validity.

6.1.2.1 Face validity

Refers to the acceptability of test items to both test users and test takers, and whether the items appear to reflect what they are supposed to measure (Rust & Golombok, 2009). With regard to parental questionnaires, feedback was gathered from professionals and parents during the test development and adaptation stage on the appropriateness of items (refer back to sections 4.2.1.2 and 4.2.2.2). In general, parents reported that the questionnaire items were clear. Some parents reported that the ARA-LUI drew their

attention to some of their children's communication skills, while some spontaneously expressed their concerns regarding the length of the ARA-LUI. However, as far as length is concerned, it is important to keep in mind that respondents in this case were participants in a research project. Possibly parents with concerns about their children's communicative abilities, whom clinicians are likely to encounter in a clinical setting would have a different motivation for filling in the questionnaire and hence be less concerned about the length of the ARA-LUI. Turning to the direct assessments (ESB, MI, SR and APS-RVT), most children engaged with the researcher during the assessment session and appeared to enjoy the tasks especially during the administration of the ESB. More details on the number of participants responding in each task are presented in section 6.3.2.

6.1.2.2 Content validity

Refers to the extent to which the content of a test is considered representative of what it was developed to measure (Rust & Golombok, 2009). The content of all the adapted and novel assessments was based on comprehensive review of the literature. In addition, the adaptation process went through careful selection of equivalent culturally appropriate items and pilot testing as described in chapter 4.

6.2.2.3 Concurrent validity

Concurrent validity refers to the extent to which the new tests correlate with existing tests that tap into the same skills (Rust & Golombok, 2009). Due to lack of standardized assessments in Saudi Arabia assessing the same skills, the validity of the adapted verbal measures (ARA-LUI and APS-RVT) was assessed by examining how they correlate with each other and with the Sentence Repetition test (see Table 7). Likewise, the validity of the nonverbal measures (ESB, MI and SCogQ) was assessed by examining their correlation with each other. Given the lack of normality in some of the measures (see Appendix H), all partial correlations controlling for age in months were performed with Bias corrected and accelerated bootstrapping (BCa) to get robust confidence intervals (CI).

Table 7: Partial correlations between verbal measures controlling for age in months

	ARA-LUI	SR Lexical	SR Grammatical	TSA
SR Lexical	.423***			
	[.201, 593]			
SR	.368***	.941***		
Grammatical	[.174, .530]	[.921, .958]		
TSA	.203 ns	.663***	.793***	
	[.026, .339]	[.562, .767]	[.696, .888]	
APS-RVT	.406***	.459***	.459***	.258*
	[.207, .556]	[.283, .612]	[.282, .636]	[.067, .536]

SR Lexical: sentence repetition lexical morpheme score; SR Grammatical: sentence repetition grammatical morpheme score; TSA: Total sentence accuracy score; ns = not significant ($p > .05$), *$p < .05$. ***$p < .0001$. BCa bootstrap 95% CIs reported in [ ].

As can be seen in Table 7, results showed moderate correlations between the ARA-LUI, APS-RVT and the lexical and grammatical morpheme scores of the Sentence Repetition with age in months controlled (SR administered for age groups 2 and 3 only). In contrast, the total sentence accuracy (TSA) showed weak correlations with the APS-RVT and weak relations with the ARA-LUI reaching significance in the robust correlations. The weak correlations between the TSA and the other verbal measures may be due to the fact that there was floor effect on the TSA (see section 6.3.2.3).

It should be noted that verbal measures used in this study assessed different aspects of language and thus high correlations between these were not expected. With regard to the nonverbal measures, the ESB and MI showed moderate significant correlations, $r = .497$, 95% BCa CI [.276, .650], $p < .001$, with age in months controlled. The ESB and the MI examined different skills of sociocognition that are theoretically related. Conversely, the SCogQ relations were non-significant with both the ESB, $r = .172$, 95% BCa CI [-.103, .301], $p = .143$ and MI, $r = .072$, 95% BCa CI [-.129, .257], $p = .406$. This was unexpected as many questions in the SCogQ relate to skills examined in the ESB.

6.2.2.4 Construct validity

Construct validity is considered to be the primary source of evidence that a test performs as expected when measuring a trait or attribute. It is inferred from accumulated evidence that justifies test interpretation and use (Rust & Golombok, 2009). One source of evidence in judging construct validity is homogeneity within the test which can be examined by intercorrelations between test subscales or tasks.

Intercorrelations between subscales of key measures

Arabic Research Adaptation of the LUI (ARA-LUI)

The ARA-LUI resembled the LUI in the number of main parts, subscales, and items. It consisted of 14 subscales divided into three parts (how your child communicates with gestures, how your child communicates with words, and your child's longer sentences) with 12 subscales scored numerically. The total score is computed from 10 subscales in part 2 and 3. Intercorrelations between the twelve scored subscales of the ARA-LUI were examined controlling for children's age in months and compared to the intercorrelations between the subscales of the original LUI reported in parentheses (see Table 8).

Table 8: Intercorrelations for all the scored subscales on the ARA-LUI controlling for children's age in months (N= 134) and correlation coefficients for the LUI (N=177, O'Neill, 2007) in parentheses

	A	B	C	D	F	G	H	I	J	K	M
B	.320*** [.277**]										
C	-.178* [-.165*]	ns [ns]									
D	ns [ns]	ns [ns]	.580*** [.787**]								
F	ns [-.155]*	ns [ns]	.637*** [.692**]	.590*** [.658**]							
G	ns [-.184*]	ns [ns]	.663*** [.704**]	.582*** [.675**]	.633*** [.740**]						
H	ns [-.260**]	ns [ns]	.679*** [.693**]	.539*** [.697**]	.596*** [.732**]	.746*** [.813**]					
I	ns [-.197**]	ns [ns]	.579*** [.617**]	.458*** [.642**]	.533*** [.651**]	.598*** [.684**]	.724*** [.844**]				
J	ns [ns]	ns [ns]	.177* [.236**]	.207* [.237**]	.252** [.392**]	ns [.421**]	.336*** [.469**]	.325*** [.459**]			
K	ns [-.250**]	ns [ns]	.346*** [.489**]	.369*** [.497**]	.365*** [.536**]	.411*** [.558**]	.570*** [.656**]	.466*** [.604**]	.346*** [.592**]		
M	ns [-.191*]	ns [ns]	.459*** [.512**]	.470*** [.546**]	.460*** [.615**]	.519*** [.645**]	.717*** [.771**]	.614*** [.746**]	.363*** [.461**]	.650*** [.698**]	
N	ns [-.189*]	ns [-.187*]	.390*** [.335**]	.362*** [.322**]	.341*** [.481**]	.446*** [.492**]	.652*** [.649**]	.487*** [.572**]	.401*** [.509**]	.657*** [.658**]	.733*** [.765**]

A: How your child uses gestures to ask for something; B: How your child uses gestures to get you to notice something; C: Types of words your child uses; D: Your child's requests for help; F: How your child uses words to get you to notice something; G: Your child's questions and comments about things; H: Your child's questions and comments about themselves or other people; I: Your child's use of words in activities with others; J: Teasing and your child's sense of humor; K: Your child's interest in words and language; M: How your child adapts conversation to other people; N: How your child is building longer sentences and stories; ns = not significant ($p > .05$), *$p < .05$. **$p < .01$, ***$p < .0001$, Correlation coefficients for the LUI (O'Neill, 2007) are in [ ].

As can be seen in Table 8, for the ARA-LUI, a significant moderate correlation was found between the first two subscales (A, B) which focus on assessing gestures (part 1) and a negative relationship between subscales A and C. On the other hand, relations of subscales A and B with all the other subscales were non-significant. With regard to the 10 remaining subscales that form parts 2 and 3, the majority of correlations between subscales were significant, with most relations (84%) ranging from $r = .33$ to $r = .68$. This pattern of

relations between subscales is similar to those reported for the LUI subscales (O'Neill, 2007) with few differences observed. For example, correlations between the subscales G and J of the ARA-LUI were non-significant, while O'Neill reported significant moderate correlations between those subscales. In addition, while most relations between subscale A and other subscales of the ARA-LUI were non-significant, O'Neill results showed mostly weak negative correlations between subscale A and other subscales. These inconsistencies might be attributed to differences in the sample, as the sample size in O'Neill's study was larger (177 children) and included a wider age range (18 months to 47 months). In addition, there might be demographic differences in the nature of the sample and SES distribution. The large similarity between results supports the evidence of validity of ARA-LUI.

Early Sociocognitive Battery (ESB)

Intercorrelations among the three tasks of the ESB (social responsiveness, joint attention and symbolic comprehension) controlling for children's age in months were measured. Results showed significant weak to moderate correlations between the three tasks (see section 6.3.2.1).

Motor Imitation (MI)

The MI test consisted of two subtasks: posture imitation and gesture imitation. The correlation between posture and gesture imitation was measured controlling for children's age in months. Strong correlations were found between posture and gesture imitation tasks (see section 6.3.2.2).

6.1.3 Summary

Good levels of reliability were found across all measures suggesting that they were fit for purpose. In addition, measures of validity support the validity of the majority of tasks. Concurrent validity was assessed by examining how the verbal measures correlated with each other. Results showed that all verbal measures correlated with each other. The majority of correlations between verbal measures were moderate. Turning to nonverbal measures, results showed moderate correlations between the ESB and MI, but unexpectedly, correlations were non-significant between the SCogQ and both the ESB and the MI. The SCogQ is a parental questionnaire that was designed to assess a number of sociocognitive skills most of which were also assessed using the direct child measures

(ESB and MI test). This finding raises questions about the validity of this measure and calls for further examination of the usefulness of this measure.

Newly developed instruments are usually evaluated by comparing their accuracy to an accepted diagnostic measure (Dollaghan, 2004). However, this type of external validation was not possible in this study as no such gold standard measures exist in Saudi Arabic.

6.2 Gender

Several studies have reported slight gender differences in early language abilities favoring girls (Bornstein, Hahn, & Haynes, 2004; Fenson et al., 1994; Wallentin, 2009). In addition, gender differences led to gender-related norms being developed for a number of early language assessments including the LUI (O'Neill, 2009). Thus, it was of interest to know whether there were gender differences in the performance of children on the different measures in this study. Mean values, medians and standard deviations of girls and boys on the different assessments are presented in Table 10. Since there was no significant difference in the age of girls and boys (see Table 9), age was not taken into account in examining the gender effect.

Table 9: Distribution of participants according to age group and gender

	Girls	**Mean age in months**	**Median age in months**	**SD**	**Boys**	**Mean age in months**	**Median age in months**	**SD**
Age group 1 (2;0-2;5)	25	26.72	27.00	1.88	26	26.12	26.00	1.73
Age group 2 (2;6-2;11)	28	32.25	32.00	1.55	27	33.04	34.00	1.77
Age group 3 (3;0-3;5)	30	39.33	40	1.92	25	38.88	39	2.24
Total	83	33.14	32	5.48	78	32.60	33.50	5.54

Table 10: Descriptive statistics of girls and boys on all tasks

	Gender	No	Mean	Median	SD
ARA-LUI	Girl	66	104.71	106.50	32.32
	Boy	68	101.00	103.00	31.69
SCogQ	Girl	66	27.50	27.50	3.31
	Boy	69	27.45	28.00	3.71
ESB	Girl	83	36.64	37.00	6.76
	Boy	78	34.06	35.00	6.25
MI	Girl	83	15.16	17.00	5.84
	Boy	78	14.71	15.00	5.07
SR Lexical	Girl	58	25.97	25.50	14.50
	Boy	52	22.44	21.00	12.80
SR Grammatical	Girl	58	39.66	34.00	29.56
	Boy	52	30.87	24.00	23.04
TSA	Girl	58	3.90	.50	6.80
	Boy	52	2.00	.00	3.66
APS-RVT	Girl	82	23.48	21.50	12.70
	Boy	78	20.27	20.00	10.63

To investigate the effect of gender, data sets were first checked for assumptions of normality and homogeneity. Assumptions were violated in most data sets (see results of Kolmogorov-Smirnov test of normality in Appendix H). Consequently, the non-parametric Mann-Whitney test was used to explore differences in performance between girls and boys on all measures. In addition, effect size which shows the actual degree of difference between the scores of different groups was calculated. Pearson's *r* effect size was calculated by converting z-score using the following formula: $r = \frac{\sqrt{z}}{n}$

Pearson's *r* effect size was interpreted using the following guidelines (Field, 2013):

Small 0.1

Medium 0.3

Large 0.5

Results showed no significant differences in the performance of girls and boys on any measure except the ESB on which girls performed significantly better than boys, with weak effect size (U = 2440; N_1 = 83, N_2 = 78; p = .007, r =.213). The non-significant difference between boys and girls on the ARA-LUI is in contrast to the norming sample of the LUI (O'Neill, 2009) in which a significant gender difference led to the development of separate norms for boys and girls. However, it is important to note that no effect of gender

on performance on the LUI was found in an earlier study with a smaller sample size (O'Neill, 2007).

The finding of a gender effect on the ESB is in line with the findings of Chiat and Roy (2006) who reported a significant small effect of gender on the performance of girls and boys in their sample of young clinically referred children (P. Roy, personal communication, August, 27, 2014). The significant gender effect was further investigated by performing three separate Mann-Whitney tests to investigate whether the effect of gender was found in the three age groups. Significance level was adjusted to .016 due to multiple comparisons. Results indicated that the gender effect was significant only in the third age group (U = 175.5; N_1= 30, N_2 = 25; p = .001). This finding is also illustrated in box plots (see Figure 2).

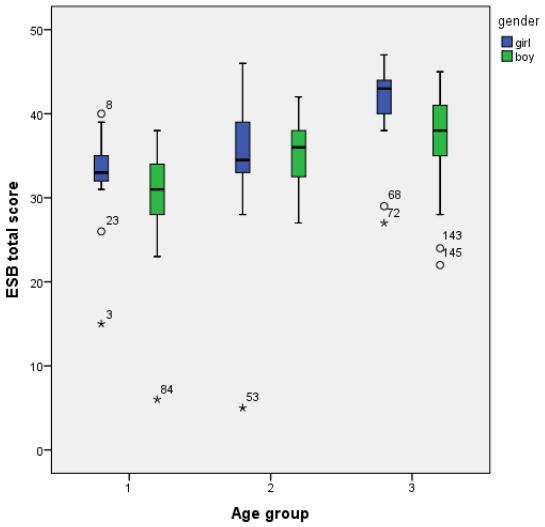


Figure 2: Box plots showing scores of children on the ESB according to age group and gender

Since the gender effect was only found in the ESB and this difference was limited to the oldest age group, a decision was taken not to include gender in subsequent analyses. All further analyses were based on 6-month age bands.

6.3 Descriptive and Inferential Statistics

The main aim of this study was to develop a battery of language and nonverbal sociocognitive measures to assess early skills in young Saudi children. Distribution of scores across the three age groups on the different measures will be described. In all cases, means, medians, standard deviations (SD) and ranges are presented. These are supported

by box plots. Box plots are well suited for illustrating the distribution of scores in newly developed or adapted measures. They show how data are spread around the median. The box shows the middle 50% of scores. The thick horizontal line in the middle of the box represents the value of the median and the hinges at either side of the box represents the interquartile ranges. The bottom hinge shows the range between which the lower 25% of scores fall and the top hinge represents the top 75% data point. The whiskers at either side of the box approximately symbolize the lowest and highest scores. Outliers are represented by circles above or below box plots and extreme outliers are represented by an asterisk. The number next to the circle or asterisk relates to the row number in the data editor (Mayers, 2013).

Differences in performance across age groups were then examined. Assumptions of normality were violated in most data-sets. This was evident in the distribution of scores (see Appendix H) and further confirmed by the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. Based on the common suggestion, in groups larger than 50 the Kolmogorov-Smirnov test was used, whereas the Shapiro-Wilk test was used for samples smaller than 50 (Mayers, 2013). Results of tests of normality are found in Appendix H. Normally distributed data is one of the assumptions of analysis of variance (ANOVA). However, ANOVA is robust to modest violations of parametric requirements (Mayers, 2013). Therefore, an independent one-way ANOVA was used to investigate the effect of age across different measures. To ensure that non-normality did not affect the results, the analyses were repeated using non-parametric Kruskal-Wallis and Mann-Whitney tests as reported in Appendix I. In cases where the data were particularly poorly distributed the non-parametric Kruskal-Wallis or Mann-Whitney was used. Cohen's *f* effect size was calculated using the G* Power software and interpreted for one-way ANOVA based on Cohen's conventions (1992) as:

Small .10

Medium .25

Large .40

On the other hand, Pearson's *r* effect size was reported for non-parametric tests.

6.3.1 Parental questionnaires

The first part of the assessment battery used two parental questionnaires, the Arabic Research Adaptation of the LUI and the Sociocognitive Questionnaire.

6.3.1.1 Arabic Research Adaptation of the LUI (ARA-LUI)

The ARA-LUI is a parental questionnaire that focused on pragmatic language development and consisted of 3 major parts, with a total of 180 questions. However, the total score is computed from parts 2 and 3 only. The maximum raw score was 161 (see section 5.3.1.1) and low scores are indicative of low performance. Descriptive statistics of the three age groups on the ARA-LUI are shown in Table 11 and further illustrated in box plots (Figure 3).

Table 11: Descriptive statistics of the three age groups on the ARA-LUI

	ARA-LUI (Maximum Score = 161)					
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	46	82.76	84	28.76	7	138
2 (2;6-2;11)	46	104.28	109	28.46	4	153
3 (3;0-3;5)	42	123.21	128	25.14	58	157

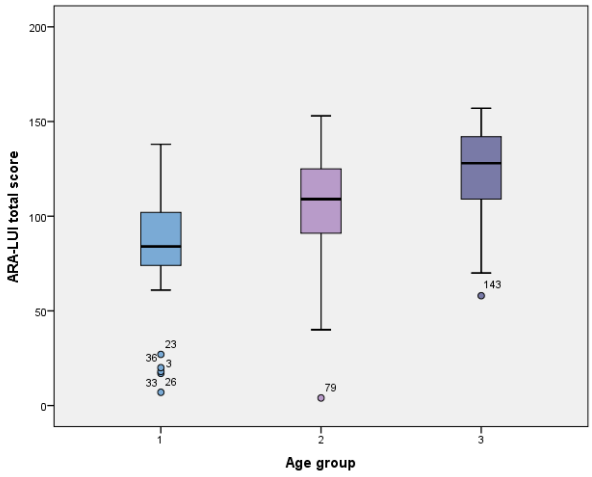


Figure 3: Box plots of the ARA-LUI scores according to age group

Fully complete ARA-LUI were available for 134 children (83.23%) of the 161 children in the sample. As can be seen in Table 11 missing data were equally distributed across the three age groups. Of the remaining 27 for whom a score on the ARA-LUI was not available, the majority (24) were due to parents not returning the questionnaire and the remaining three were excluded as they were minimally completed. The maximum score achieved was 157. As can be seen in the box plots (Figure 3) scores of children in the youngest age group appear to be positively skewed whereas scores of children in age

group 3 appear to be negatively skewed. The greatest spread of scores was found in age group 2. Differences in performance across age groups were evident in the box plots. This observation was further investigated using one-way ANOVA. The effect of age was significant F (2,131) = 23.73, $p < .001$. *Post-hoc* analysis was performed using Gabriel's method since the sample sizes were different. The Gabriel's *post-hoc* test revealed a significant difference in scores of children in age group 1 as compared to scores of children in age group 2 ($p = .001$) and age group 3 ($p < .001$), and a significant difference between scores of children in age group 2 and age group 3 ($p = .005$). The effect size for these age group differences was large, $f = 0.51$. Thus, results showed that the ARA-LUI like the original LUI (O'Neill, 2007) was sensitive to age. The range of scores of children on the ARA-LUI in the current study was compared to the Canadian normative sample. The median score of children in each age group in the Saudi sample was compared to the average of the 50th percentile score of boys and girls in the equivalent age range in the Canadian sample. Results showed the scores in the current study tended to be lower than scores for Canadian children on the LUI (see Appendix J). One possible reason for this finding is the differences in the inclusion criteria between the two studies as children with diagnosed speech or language problems were excluded in the LUI study (O'Neill, 2009).

6.3.1.2 Sociocognitive Questionnaire (SCogQ)

The SCogQ is a novel questionnaire consisting of 18 questions covering early nonverbal social and cognitive skills such as joint attention, pretend play, empathy and imitation. The maximum possible score was 36. Low scores are indicative of a more problematic outcome whereas higher score are indicative of a more positive outcome (see section 5.3.1.2). Table 12 shows the descriptive statistics for the SCogQ and Figure 4 presents the box plots for the three age groups.

Table 12: Descriptive statistics of the three age groups on the SCogQ

		SCogQ (Maximum Score = 36)				
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	45	27.64	28	3.70	18	33
2 (2;6-2;11)	47	27.91	27	3.16	20	35
3 (3;0-3;5)	43	27.91	29	3.65	17	32

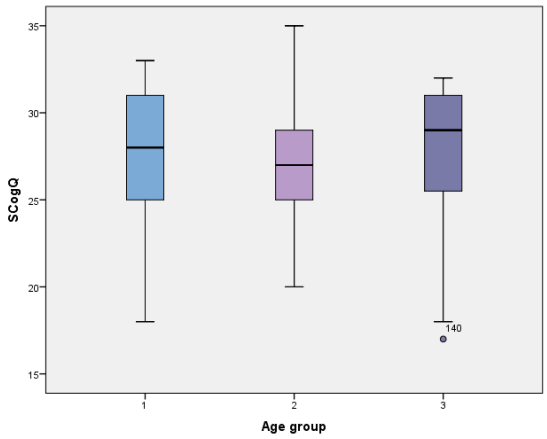


Figure 4: Box plots of the SCogQ scores according to age group

Like the ARA-LUI, there were missing data for children on the SCogQ mainly due to non-return of the questionnaire. The total number of returned questionnaires was 135. Once again, the distribution of missing data across the three age groups was relatively equal. None of the children achieved the maximum score. Mean and median scores were very similar across age groups and one-way ANOVA confirmed no significant difference between scores of the three age groups, F (2, 132) = .979, p = .378, f = 0.12. Thus, in contrast to the ARA-LUI, the SCogQ was not sensitive to age. The items included in the questionnaire assess skills that were reported to relate to language abilities most of which emerge before 2 years of age (Carpenter et al., 1998).

6.3.2 Direct measures

Four direct assessments were administered, the Early Sociocognitive Battery (ESB), the Motor Imitation test (MI), the Sentence Repetition test (SR) and the Arabic Preschool-Receptive Vocabulary Test (APS-RVT), although the SR test was not administered to the youngest age group (2;0-2;5).

6.3.2.1 Early Sociocognitive Battery (ESB)

The ESB included three tasks: social responsiveness, joint attention and symbolic comprehension. The social responsiveness task measures the child's responses to feelings expressed by the examiner by looks to the examiner's face. In the joint attention task, the child's gaze alternation, gaze-following and point-following are measured. The symbolic

comprehension assesses the child's understanding of different levels of symbolic comprehension (see section 5.3.2.1).

ESB tasks

Table 13 shows the descriptive statistics for the three tasks of the ESB according to age group and Figure 5 illustrates the box plots of the scores.

Table 13: Descriptive statistics of the three age groups on the three tasks of the ESB

Social Responsiveness (Maximum Score = 12)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	9.94	10	1.87	4	12
2 (2;6-2;11)	55	10.07	10	1.67	4	12
3 (3;0-3;5)	55	10.40	11	1.51	7	12
Joint Attention (Maximum Score = 18)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	14.45	15	2.31	2	18
2 (2;6-2;11)	55	14.73	15	2.68	1	18
3 (3;0-3;5)	55	15.89	16	1.81	10	18
Symbolic Comprehension (Maximum Score = 18)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	6.92	7	3.14	0	13
2 (2;6-2;11)	55	10.22	11	3.64	0	17
3 (3;0-3;5)	55	13.25	15	3.48	3	18

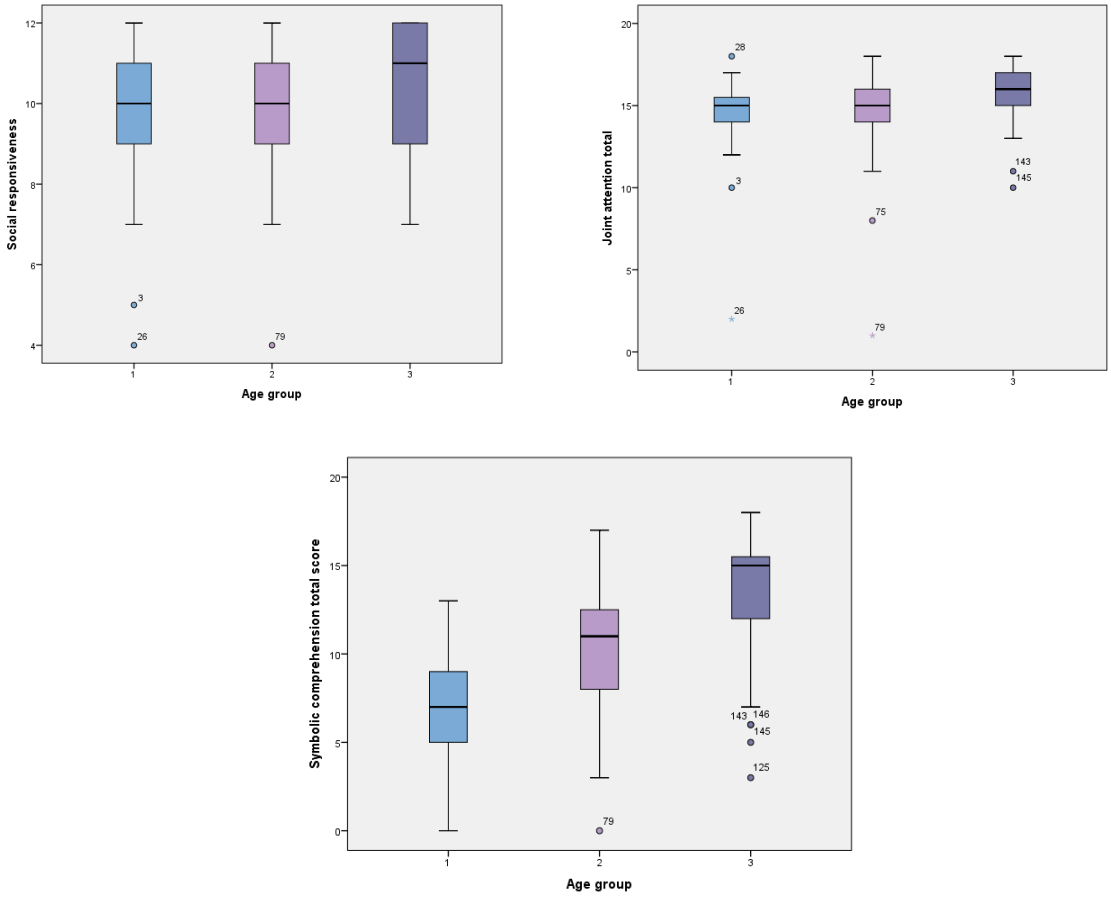


Figure 5: Box plots for scores on the three tasks of the ESB according to age group

As shown in Table 13 and Figure 5, most children scored highly on the social responsiveness task, with a number of children in the three age groups achieving the maximum possible score (21.57% of age group 1, 18.18% of age group 2, and 27.27% of age group 3). Fewer children reached the maximum possible score on the joint attention (1.96% of age group 1, 5.45% of age group 2, and 16.36% of age group 3) and only 1 child in the third age group (1.82%) achieved the maximum score on the symbolic comprehension task. Outliers appeared in all the tasks in different age groups with two extreme outliers in the joint attention task from the two youngest age groups. Most outliers were the same children across two or more tasks. For example, the outlier in the second age group was an outlier in all tasks. In addition, the two outliers found in the youngest age group in the social responsiveness task were also outliers in the joint attention task. However, on the symbolic comprehension task, children who scored 0 represented the lower end of the range in age group 1, thus the two low scoring children on social

responsiveness and joint attention from the youngest age group did not emerge as outliers in the symbolic comprehension task.

Among the tasks of the ESB the most marked difference in performance across age groups was found in the symbolic comprehension task. The effect of age in the three tasks was further investigated using one-way ANOVA. Results showed that the age effect was significant for both the joint attention F (2,158) = 5.98, $p < .01$ and symbolic comprehension F (2,158) = 45.04, $p < .001$ tasks. On the other hand, there was no significant age difference on the social responsiveness task F (2,158) = 1.06, $p = .35$, $f = 0.12$. Gabriel's *Post-hoc* analysis showed that on the joint attention task age group 3 performed significantly better than both age group 1 ($p = .026$) and age group 2 ($p = .004$). On the other hand, there was no significant difference in performance between age groups 1 and 2 on the joint attention task. With regard to the symbolic comprehension task, Gabriel's *Post-hoc* showed that there was a significant difference in performance between all three age groups ($p < .001$). These results are in line with Chiat and Roy's (2006) results as only the joint attention and symbolic comprehension tasks were found to be age-sensitive in their typically developing sample.

Correlations between the three tasks were examined controlling for children's age in months with bias accelerated bootstrapping. As seen in Table 14, moderate correlations were found between the joint attention and both the social responsiveness and symbolic comprehension. In addition, weak significant correlations between the social responsiveness and symbolic comprehension were found although it must be noted that the 95% bias corrected bootstrapped CI for the social responsiveness and joint attention correlation ranged from -.006 to .521, thus crossing the zero indicating that the association is not significant and less robust.

Table 14: Intercorrelations between the three tasks of the ESB controlling for age in months

	Social responsiveness	**Joint Attention**
Joint attention	.307***	
	[-.006, .521]	
Symbolic comprehension	.259**	.459***
	[.076, .421]	[.299, .590]

**$p < .01$, ***$p < .001$, BCa bootstrap 95% CIs reported in [ ].

The three tasks of the ESB tap into different skills of sociocognition. The symbolic comprehension task correlated with both the social responsiveness and joint attention with robust confidence intervals. In addition, the three tasks showed large significant correlations with the ESB total score (sum of scores of the three tasks) when age in months was partialled out; social responsiveness r =. 586, 95% BCa CI [.414, .703], p < .001, joint attention r = .777, 95% BCa CI [.612, .865], p < .001, symbolic comprehension r = .853, 95% BCa CI [.809, .894], p < .001. The use of a total score would increase the breadth of the measure and the likelihood of increasing its predictive value. Moreover, it has been shown that measurement error is reduced when an aggregate measure is used (Rushton, Brainerd, & Pressley, 1983). The total score of the ESB was used in all further analyses. Inspecting the box plots of the ESB total score (Figure 6) further supports the validity of using a composite score, as children who performed low and appeared as outliers in the tasks of the ESB emerged as outliers when the composite score was used.

ESB total score

The maximum for the total score of the ESB was 48. Table 15 presents the descriptive statistics for the total score of the ESB according to age group and Figure 6 illustrates the box plots of the scores.

Table 15: Descriptive statistics for the ESB total score for the three age groups

	ESB Total Score (Maximum Score = 48)					
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	31.31	32	4.55	6	40
2 (2;6-2;11)	55	35.02	36	4.17	5	46
3 (3;0-3;5)	55	39.55	41	5.63	22	47

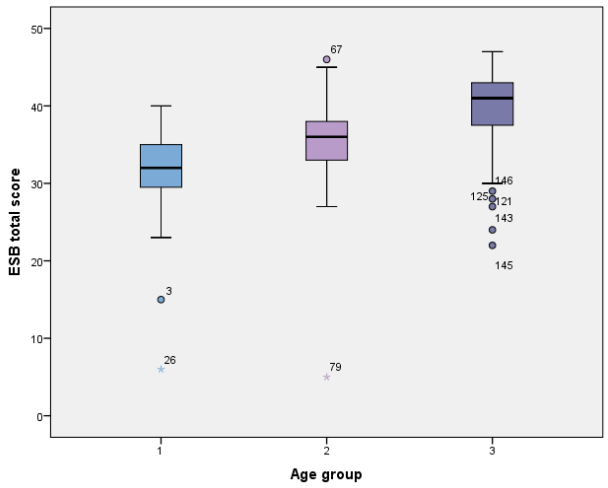


Figure 6: Box plots of the ESB total scores according to age group

As seen in Figure 6, outliers appeared in all three age groups, with two extreme outliers in age groups 1 and 2. Most of these children were the same participants who emerged as outliers on the individual tasks of the ESB.

With regard to differences in performance across age groups, there appears to be an increase in the total score of the ESB with age although none of the children achieved the maximum score. One-way ANOVA showed that the effect of age was significant, $F(2,158) = 27.305$, $p < .001$. Gabriel *Post-hoc* tests further showed that age group 1 performed significantly different from age groups 2 ($p = .003$) and 3 ($p < .001$). In addition, performance of age group 2 was significantly different from age group 3 ($p < .001$). The effect size was large, $f = 0.51$. Thus, the ESB was shown to be an age sensitive measure of social cognition.

6.3.2.2 Motor Imitation (MI)

The Motor Imitation test included the two tasks of posture imitation and gesture imitation. In the posture imitation task, children imitated 6 different hand movements that did not have semantic or conventional communicative meaning associated with it, whereas in the gesture task they imitated 6 different gestures that symbolized a referent. Children's responses were scored on a 3 point scale based on the accuracy of their responses: an accurate response was awarded 2 points, a partially accurate response one point and no points for failure. The total maximum score for the posture and gesture tasks was 24 (see section 5.3.2.2). In the MI test, a number of children did not respond either to some or all of the items of the task. Non-responses were scored as 0, as refusing to respond might be an indication of inability rather than unwillingness to imitate, these children were not

excluded to avoid losing important information about their imitation performance (Dohmen, Chiat & Roy, 2013).

Motor Imitation Tasks (Posture and Gesture Imitation)

Table 16: Descriptive statistics for the subtasks of the Motor Imitation

Posture Imitation (Maximum Score = 12)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	6.76	8	3.40	0	12
2 (2;6-2;11)	55	8.78	9	2.45	0	12
3 (3;0-3;5)	55	10.00	11	2.38	0	12
Gesture Imitation (Maximum Score = 12)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	4.82	5	2.88	0	9
2 (2;6-2;11)	55	6.53	7	2.24	0	10
3 (3;0-3;5)	55	7.67	8	2.55	0	12

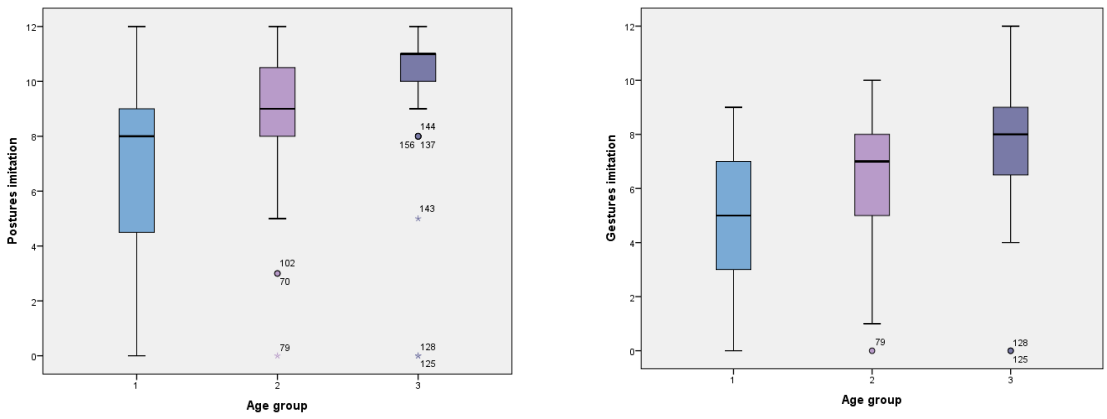


Figure 7: Box plots of the scores on the subtasks of the Motor Imitation according to age group

As observed in Table 16 and Figure 7, performance on the two tasks of the Motor Imitation test increased with age. A one-way ANOVA was performed with Welch's *F* adjustment as homogeneity of variance was not assumed. Results showed that the effect of age was significant for the postures, F (2,158) = 18.23, $p < .001$.Welch's *F* adjustment showed that the violations of homogeneity of variance had not impacted the observed outcome. Games-Howell *Post-hoc* showed that the three age groups performed significantly differently from each other (age groups 1and 2, $p = .002$, age groups 1and 3, $p < .001$, age groups 2 and 3, $p = .022$). Likewise, one-way ANOVA showed a significant age effect for the gesture tasks**,** F (2,158) = 16.47, $p < .001$. In addition, Gabriel's *Post-hoc* showed that age group 1 performed significantly differently from both age group 2 ($p = .002$) and age group 3 ($p < .001$), whereas no significant difference was found between age groups 2 and 3. Comparing performance of all children on the posture and gesture tasks showed that in general children performed better on the posture task with some children in the three age groups achieving the maximum score (12) and the median score in the third age group on the posture task was 11 showing ceiling effect of this task in the third age group. This might be due to the fact that the posture task being physically less demanding than the gesture task. All children who appeared as outliers in the gesture task appeared as extreme outliers in the posture task. Six more children emerged as outliers in the posture task (age group 2 = 2, age group 3 = 4). Unlike the gesture task, scores of the majority of children in the posture task in age group 3 were spread in a narrower range with most children scoring highly, and the few children with low performance appearing as outliers. In addition, the widest range of scores was in the youngest age group on the posture task.

Correlation between the two tasks of the Motor Imitation test was examined controlling for age in months. Results showed that two tasks were highly correlated, $r = .665$, 95% BCa CI [.524, .781], $p < .001$, which validates the use for the total score of the Motor Imitation test.

Motor Imitation Total Score

The maximum score for the Motor Imitation total score was 24 which is the sum of the posture imitation and gesture imitation total scores.

Table 17: Descriptive statistics for the Motor Imitation total score

Motor Imitation Total (Maximum score = 24)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	11.59	13	5.90	0	21
2 (2;6-2;11)	55	15.31	15	4.17	0	22
3 (3;0-3;5)	55	17.67	19	4.54	0	24

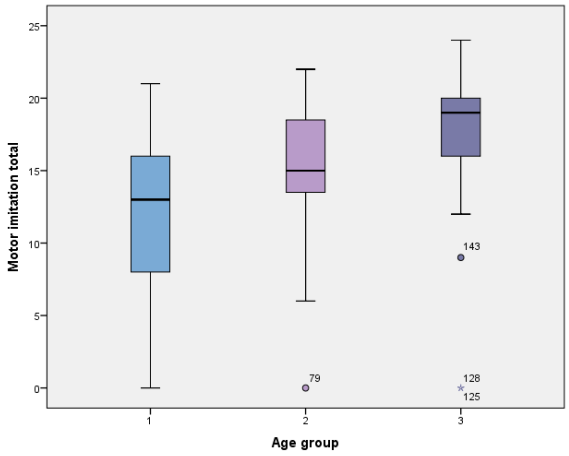


Figure 8: Box plots of the total scores for the Motor Imitation according to age group

Nine children did not respond or refused to participate in the Motor Imitation task, of whom six were from the youngest age group (11.76% of age group 1), 1 from age group 2 (1.82%) and 2 from age group 3 (3.64%). These children scored 0. Two more children from the youngest age group imitated one or more of the items in the posture task but refused to imitate any item in the gesture task. Results are shown in Table 17 and box plots in Figure 8. Children's scores increased and the range of scores narrowed with age, with one child in the third age group achieving the maximum score (24). These observations were further investigated using one-way ANOVA with Welch's *F* adjustment as homogeneity of variance was not assumed. Analysis showed a significant effect of age, F (2, 158) = 20.659, $p < .001$. The violation in homogeneity of variance had no impact on observed outcome as shown in Welch's *F* adjustment. Games Howell *Post-hoc* was used since equality of variance was not assumed. Results showed a significant difference between age group 1 in comparison to age group 2 ($p = .001$) and age group 3 ($p < .001$).

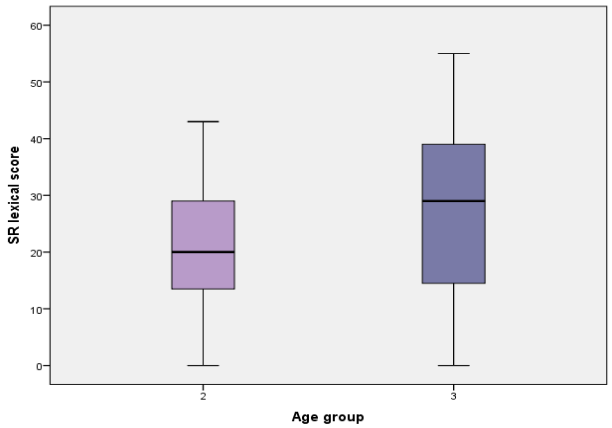
A significant difference was also found between performance of children in age group 2 and age group 3 (p = .015). Effect size was large, f = 0.45. Thus, children's performance on the Motor Imitation test improved significantly with age. Interestingly, comparing the present results to the findings of Dohmen, Chiat and Roy (2013) showed that in both studies the greatest percentage of refusals was found in the youngest age group.

6.3.2.3 Sentence Repetition (SR)

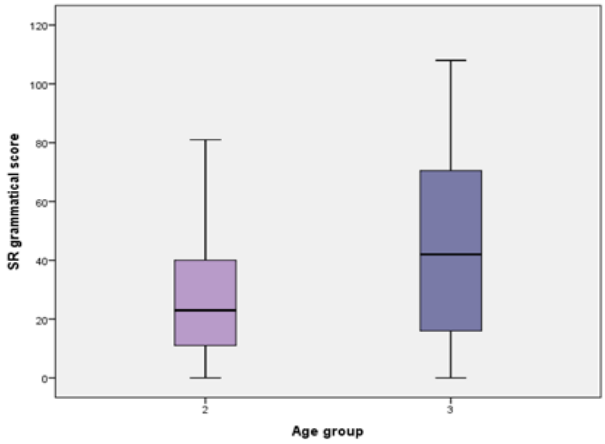
The Sentence Repetition test was administered only to age groups 2 and 3. This task assesses the children's ability to repeat 14 sentences produced one at a time by the examiner. Three scores were obtained for children on the Sentence Repetition task: the lexical morpheme score (SR Lexical) maximum score = 56, the grammatical morpheme score (SR Grammatical) maximum score = 117, and the total sentence accuracy (TSA) maximum score = 42 (see section 5.3.2.3 and Appendix F). Children who refused to participate or did not respond in the Sentence Repetition test were scored 0. Table 18 presents the descriptive statistics for the Sentence Repetition test and Figure 9 shows the box plots for the three scores.

Table 18: Descriptive statistics for the three scores of the Sentence Repetition

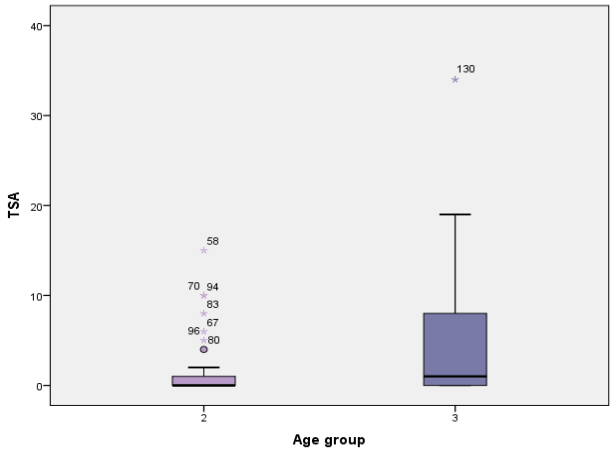
SR Lexical (Maximum score = 56)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
2 (2;6-2;11)	55	20.53	20	11.70	0	43
3 (3;0-3;5)	55	28.07	29	14.74	0	55
SR Grammatical (Maximum score = 117)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
2 (2;6-2;11)	55	27.16	23	20.30	0	81
3 (3;0-3;5)	55	43.84	42	30.15	0	108
TSA (Maximum score = 42)						
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
2 (2;6-2;11)	55	1.29	.00	3.04	0	15
3 (3;0-3;5)	55	4.71	1	6.94	0	34



- SR Lexical Morpheme Score (Maximum Score = 56)



- SR Grammatical Morpheme Score (Maximum Score = 117)



- Total Sentence Accuracy (Maximum Score = 42)

Figure 9: Box plots for the three scores of the Sentence Repetition according to age group

Nine children (age group 2 = 5, age group 3 = 4) did not respond or refused to participate. Among those children, one child from age group 2 and two children from age group 3 were also scored 0 on the Motor Imitation test due to non-compliance. Inspecting box plots Figure 9 show that for both the lexical and grammatical morpheme scores there was a wider range of scores for age group 3 extending higher, as compared to age group 2. No outliers appeared in any age group for either the lexical or grammatical score and none of

the children achieved the maximum score on any of the three scores. There was a floor effect for age group 2 on the TSA with the median score = 0. Most children (72.72%) in age group 2 scored 0 on the TSA as compared to 36.36% in age group 3. To investigate effects of age on lexical and grammatical scores, one-way ANOVA with Welch's *F* adjustment was used as there were violations of homogeneity of variance. Results showed that the effect of age was significant with a medium effect size for both the lexical score F (1, 108) = 8.846, $p = .004$, $f = .27$ and the grammatical score F (1, 108) = 11.573, $p = .001$, $f = .31$. Welch's *F* adjustment showed that the violations of homogeneity of variance had no impact on observed outcomes. The effect of age on the TSA was investigated using the non-parametric Mann-Whitney as the scores were poorly distributed. The effect of age was significant, with medium effect size (U = 926.0, $N_1 = 55$; $N_2 = 55$; $p < .001$, $r = .37$). Thus, children showed a significant improvement in performance on all the scores of the Sentence Repetition test with age. Very similar results were found by (Wallan, Chiat & Roy, 2011) with only slight differences in mean scores (see Appendix J) which may be due to differences in the recruitment and inclusion criteria between the two studies. In Wallan and colleagues' study mean scores were calculated for the group of typically developing children which was defined as children with no concerns about their language development. None of the children in Wallan and colleagues' study refused to participate in the Sentence Repetition test.

6.3.2.4 Arabic Preschool-Receptive Vocabulary Test (APS-RVT)

The APS-RVT included 84 items divided into 7 groups each containing 12 items. Once the child made 8 errors in one group testing was completed for that group but not continued to the following group. The maximum possible score was 84 (see section 5.3.2.4). Table 19 shows the descriptive statistics for the APS-RVT and Figure 10 illustrates the box plots of the scores of children on the APS-RVT according to age group.

Table 19: Descriptive statistics of the three age groups on the APS-RVT

		APS-RVT (Maximum score = 84)				
Age Group	**n**	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
1 (2;0-2;5)	51	13.02	14	6.75	0	32
2 (2;6-2;11)	55	22.31	21	9.92	0	54
3 (3;0-3;5)	54	29.91	28.50	11.57	0	53

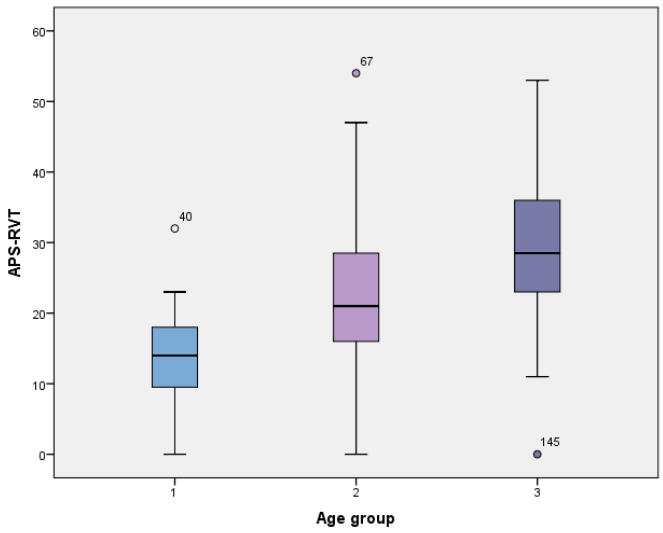


Figure 10: Box plots of the scores of the APS-RVT according to age group

One child from the oldest age group refused to complete the assessment session and did not participate in the APS-RVT. This child was excluded from this task and not represented in the box plot. As can be seen in Table 19, the minimum score for children from the three age groups on the APS-RVT was 0. The child from the oldest age group who scored 0 showed a difficulty in following instructions when compared to other children in the same age group. Thus, this child appeared as an outlier in the box plot Figure 10. The range of scores in the three age groups in Table 19 and Figure 10 shows that there is a trend for scores to increase with age with the greatest spread of scores found in age group 2. One-way ANOVA showed that effect of age was significant F (2, 157) = 40.067, $p < .001$. Welch's F adjustment showed that the violation in homogeneity of between group variance had no impact on the observed outcome. Games-Howell *post-hoc* analysis indicated that age group 3 performed significantly better than children in age group 2 ($p = .001$) and age group 1 ($p < .001$). Children in age group 2 also performed significantly better than children in age group 1 ($p < .001$). The effect size was large, $f = .58$. Thus, the APS-RVT was an age sensitive measure of children's receptive vocabulary.

6.3.3 Summary

The assessment battery included two parental reports (Arabic Research Adaptation of the LUI and the Sociocognitive Questionnaire) and three child measures (Early Sociocognitive Battery, Motor Imitation and Arabic Preschool Receptive Vocabulary Test) with one more task (Sentence Repetition) administered only to age groups 2 and 3.

No significant difference was found between girls and boys on any task except on the Early Sociocognitive Battery which showed a marginal advantage for the girls. Since this difference was not observed across the three age groups, the data was analyzed only based on age groups. The results showed that all the assessments were sensitive to age with the exception of the sociocognitive questionnaire (SCogQ) on which children from the three age groups performed similarly. Measures that showed an increase of raw scores with age have the potential to be valid tools for discriminating between typically developing children and children with language problems.

Among the direct child measures, children who refused to participate or did not respond were relatively rare and mainly found in the Motor Imitation and the Sentence Repetition tests. Refusals and non-responses were scored 0 based on the argument that refusal might be an indication of difficulty rather than uncooperativeness (Chiat & Roy, 2006; Dohmen, Chiat & Roy, 2013). In the Motor Imitation test most of the children who did not participate in the task came from the youngest age group, whereas on the Sentence Repetition test on which the youngest age group was not assessed, 55.56% of children who did not participate were from age group 2 and 44.44% from age group 3. All the children who did not participate in the Motor Imitation test from age group 2 and 3 did not participate in the Sentence Repetition test either. At this point, it is difficult to explain the reason for non-compliance of participants in this project. However, this matter will be further investigated in section 6.5.1.

On the Sentence Repetition test, the total sentence accuracy (TSA) showed floor effects. This rules out the possibility of children aged 2;6-3;5 falling below the normal range. Hence, the TSA score will not be included in subsequent analyses.

In general, results of the current study using same or similar measures, were in agreement with previous research (Chiat & Roy, 2006; Dohmen, Chiat & Roy, 2013; O'Neill, 2007) which adds to the validity of the measures.

6.4 Relations between Measures

Relations between measures were investigated for the whole sample using correlations controlling for age in months. Predictors of concurrent language were then investigated using regressional analyses using measures that showed significant correlations.

6.4.1 Correlations between sociocognitive skills and language skills

Assumptions of parametric correlations were not met in some data sets in terms of normality and outliers so partial correlations were performed using bias corrected and accelerated bootstrapping to get robust confidence intervals. Table 20 shows the magnitude of partial correlations between the sociocognitive skills as measured on the Early Sociocognitive Battery (ESB), Motor Imitation test (MI), and Sociocognitive Questionnaire (SCogQ) and the language skills measured on the Arabic Research Adaptation of the LUI (ARA-LUI), and the Arabic Preschool Receptive Vocabulary Test (APS-RVT). Scores for the Sentence Repetition test were not included in this analysis as this task was not administered to the youngest age group. It must be noted that a number of children had missing data on some measures especially on the parental questionnaires as completed questionnaires were available only for 134 out of the 161 participants. Missing data were deleted casewise.

Table 20: Partial correlations between the ARA-LUI, SCogQ, ESB and MI for the whole sample controlling for age in months

	ARA-LUI	**SCogQ**	**ESB**	**MI**
SCogQ	.323*** [.146, .486]			
ESB	.552*** [.330, .711]	ns		
MI	.420*** [.227, .577]	ns	.479*** [.257, .666]	
APS-RVT	.405*** [.240, .536]	ns	.523*** [.382, .690]	.344*** [.199, .486]

ARA-LUI: Arabic research adaptation of the LUI; SCogQ: Sociocognitive Questionnaire; ESB: Early Sociocognitive Battery; MI: Motor Imitation; APS-RVT: Arabic Preschool Receptive Vocabulary Test; ns = not significant ($p > .05$), ***$p < .001$, BCa bootstrap 95% CIs reported in [ ]

As shown in Table 20, the sociocognitive tasks correlated significantly with all the language measures except for the SCogQ which only showed significant moderate correlations with the ARA-LUI. The correlations between the SCogQ and ARA-LUI must be interpreted with caution as it might be a reflection of the common methods of assessment and not the skills being assessed as both were parent-completed questionnaires. Highest correlations were found between the ESB and the ARA-LUI, r = .552, 95% BCa CI [.330, .711], $p < .001$ and between the ESB and the APS-RVT, r =

.523, 95% BCa CI [.382, .690], $p < .001$. All other correlations between the ARA-LUI, ESB, MI, and APS-RVT were moderate positive correlations. Interestingly, the magnitude of correlations between the ESB and receptive and expressive language measures is similar to that found in Chiat and Roy's study with their clinically referred sample when age was controlled for (ESB and Auditory PLS $r = .63$, ESB and Expressive PLS $r = .55$) (P. Roy, personal communication, March, 20, 2015).

The next analysis examined correlations between all tasks for age groups 2 and 3. The total sentence accuracy (TSA) was not included in this analysis as descriptive statistics showed floor effects (see section 6.3.2.3). Table 21 shows the degree of correlations between all measures controlling for age in months.

Table 21: Partial correlations between all the measures for age groups 2 and 3 controlling for age in months

	ARA-LUI	**SCogQ**	**ESB**	**MI**	**SR Lexical**	**SR Grammatical**
SCogQ	.275* [.039, .467]					
ESB	.559*** [.277, .734]	ns				
MI	.413*** [.138, .616]	ns	.450*** [.099, .722]			
SR Lexical	.423*** [.205, .593]	ns	.26* [.027, .468]	.376*** [.147, .544]		
SR Grammatical	.368*** [.138, .616]	ns	.275* [.107, .456]	.310* [.120, .476]	.94*** [.921, 959]	
APS-RVT	.406*** [.192, .569]	ns	.574*** [.385, .715]	.358** [.188, .529]	.459*** [.263, .627]	.459*** [.269, .644]

ARA-LUI: Arabic Research Adaptation of the LUI; SCogQ: Sociocognitive Questionnaire; ESB: Early Sociocognitive Battery; MI: Motor Imitation; SR Lexical: sentence repetition lexical morpheme score; SR Grammatical: sentence repetition grammatical morpheme score; APS-RVT: Arabic Preschool Receptive Vocabulary Test; ns = not significant ($p > .05$), *$p < .05$, **$p < .01$, ***$p < .001$, BCa bootstrap 95% CIs reported in [ ]

Again, as can be seen in Table 21, the SCogQ showed significant correlations only with the ARA-LUI. No other significant correlations were found between the SCogQ and other

measures. With the exception of the SCogQ, all measures correlated significantly with each other. Highest correlations between tasks were again found between the ESB and the ARA-LUI and ESB and the APS-RVT, with both showing strong positive correlations. The ESB showed weak correlations with the Sentence Repetition scores (SR Lexical, SR Grammatical). The MI showed moderate correlations with the ARA-LUI, APS-RVT and the two scores of the SR.

6.4.2 Regression

The assessment battery in this project used measures that have been reported to be good predictors of children's language outcome both concurrently and longitudinally. Examining the extent to which different language and sociocognitive measures predict concurrent language as assessed on a gold standard diagnostic measure was not possible due to lack of such a measure in Saudi Arabic. However, a second aim of this project was to examine the unique contribution of the performance on the nonverbal sociocognitive measures to different aspects of children's language. To this end, sociocognitive measures were entered in multiple linear regression analyses with the three language measures (APS-RVT, ARA-LUI and SR lexical morpheme score) as outcome measures. In the first analysis, the ARA-LUI, which may be considered a general measure of expressive language, was the outcome measure. The second set of regressional analyses used the APS-RVT, which assesses linguistic comprehension of single words as an outcome measure. The final outcome measure was the lexical morpheme score of the Sentence Repetition task which draws on the child's phonological and morphosyntactic abilities (Seeff-Gabriel, Chiat & Dodd, 2010). These analyses will throw more light on the relations between different early sociocognitive and language skills and will provide more information on the usefulness of assessing these sociocognitive skills in clinical settings.

In total five regression analyses were performed, two using the whole sample and three using age groups 2 and 3 since the SR was only administered to the two oldest age groups. In all cases, only measures that showed significant correlations (at least $r = .3$) with the outcome were used as predictors in the regression model. Thus, the SCogQ was not included among the predictor variables. In addition, since SR lexical morpheme and grammatical morpheme scores were shown to highly correlate with each other, the SR grammatical morpheme score was not included in the analyses to avoid multicollinearity. As mentioned in section 6.3.2.3, the TSA showed floor effects and was also not included

in the regressional analyses. Since the three language outcome measures (ARA-LUI, APS-RVT, and SR) were shown to be age sensitive, age in months was forced into the model before entering the predictors simultaneously to investigate their contribution to the outcome. Assumptions of multiple linear regression were checked (normality, outliers, independent errors, linearity and multicollinearity). In most cases analyses revealed violations of the assumptions in terms of outliers. Therefore, regression analyses were rerun using bootstrapping bias-corrected 95% confidence intervals (1000 samples). Bootstrapping is an alternative method to parametric estimates that can overcome problems when assumptions are violated. It generates robust estimates of significance tests and confidence intervals of the model parameters (Field, 2013).

6.4.2.1 Predictors of pragmatic language

Hierarchical linear regression was used to examine the contribution of performance on the ESB, MI, and APS-RVT to children's pragmatic language score as measured by the ARA-LUI (see Table 22). The ARA-LUI mainly assesses children's pragmatic language, however, it also includes some aspects of semantics and syntax in its subscales. In addition, the content of the LUI was based on the premise that language use is determined by growth in social cognition (O'Neill, 2007). In the light of this and the sociocognitive hypothesis it was predicted that APS-RVT and both ESB and MI would predict pragmatic language scores on the ARA-LUI.

Table 22: Hierarchical linear regression analysis of ARA-LUI scores (n=133), with 95% bias corrected and accelerated confidence intervals in [ ]. Confidence intervals and standard errors based on 1000 bootstrap samples

Predictor variable	R^2	$Adj.R^2$	*F*	*p*	*Constant*	*b*	*SE B*	*β*	*p*
Step 1	.307	.302	58.14	<.001	-3.10 [-30.92, 23,67]		14.06		.826
Age in months						3.26 [2.46, 4.12]	0.42	.55	.001
Step 2	.550	.536	39.16	<.001	-17.47 [-43.09, 11,89]		14.22		.212
Age in months						0.97 [0.00, 1.97]	0.46	.17	.040
ESB						1.86 [0.86, 2.71]	0.46	.39	.001
MI						1.02 [0.13, 1.82]	0.43	.18	.020
APS-RVT						0.40 [-0.05, 0.82]	0.23	.14	.086

ΔR^2 = .243 for step 2

As shown in Table 22, age in months was entered in the first block, this explained 30.7% of the variance. In the second step, ESB, MI and APS-RVT were added to the model. At this point 55.0% of the variance (*Adj. R2* = .536) was explained by the model which was found to significantly predict outcome F (4,128) = 39.16, $p < .001$. In line with expectations, the ESB (b = 1.86 [0.86, 2.71], p = .001) and MI (b = 1.02 [0.13, 1.82], p = .020) added significantly to variance in ARA-LUI scores. On the other hand, APS-RVT did not contribute significantly.

Given that the ESB showed the strongest correlation with the ARA-LUI, the regression was re-run entering the predictors hierarchically to examine the amount of variance explained by the ESB. Once age was entered in the first step, ESB explained an additional 21.2% of the variance in the second step. In step 3, MI explained an additional 2.2%.

A second hierarchical linear regression investigated the proportion of variance in ARA-LUI explained by predictors for age groups 2 and 3 and the SR Lexical was added to the predictor variables (Table 23).

Table 23: Hierarchical linear regression analysis of ARA-LUI scores for age groups 2 and 3 (n = 87), with 95% bias corrected and accelerated confidence intervals in [ ]. Confidence intervals and standard errors based on 1000 bootstrap samples

Predictor variable	R^2	$Adj.R^2$	*F*	*p*	*Constant*	*b*	*SE B*	*β*	*p*
Step 1	.145	.135	14.40	<.001	11.22 [-45.15, 62.41]		26.73		.668
Age in months						2.86 [1.52, 4.35]	0.72	.38	.001
Step 2	.491	.460	15.65	<.001	-5.34 [-54.02, 53.44]		23.85		.830
Age in months						.47 [-0.83, 1.85]	0.67	.06	.487
ESB						2.10 [0.98, 3.13]	0.57	.47	.001
MI						.73 [-0.54, 1.73]	0.64	.11	.226
SR Lexical						.57 [0.12, 0.99]	0.22	.26	.004
APS-RVT						-.04 [-0.60, 0.45]	0.27	-.02	.892

ΔR^2 = .347 for step 2

As can be seen in Table 23 , age in months was entered in the first step and found to account for 14.5% of the variance in ARA-LUI scores. When the other predictors (ESB, MI, SR Lexical, and APS-RVT) were added in the second stage, 49.1% of the variance (*Adj*. R^2 =.46) was explained by the model which was found to significantly predict outcome F (5, 81) = 15.65, $p < .001$. ESB (b = 2.10 [0.98, 3.13], p = .001) and SR Lexical (b = 0.57 [.12, .99], p = .004) contributed significantly to the model. On the other hand, MI and APS-RVT did not contribute significantly. Thus, in contrast to the results of the first regression, MI when used in combination with the two predictors ESB and APS-RVT did not contribute significantly to the model once SR Lexical was added as a predictor. However, it is important to remember that the two models differed in the age range of children included. To check if the different findings were as result of the age difference, the regression analysis for age groups 2 and 3 using the ARA-LUI as outcome was rerun without the SR Lexical as a predictor. Results showed again that MI did not make a significant contribution to the model, indicating that MI contributed significantly to the ARA-LUI scores only when the youngest age group (2;0-2;5) was included.

6.4.2.2 Predictors of receptive vocabulary

Hierarchical regression analyses were conducted to investigate the contribution of the ESB, MI and SR to APS-RVT scores (see Table 24). Analysis was first performed using the whole sample. Based on the sociocognitive hypothesis (Chiat & Roy, 2008, 2013), it was expected that the APS-RVT scores would be predicted by the ESB.

Table 24: Hierarchical linear regression analysis of APS-RVT scores (n = 160), with 95% bias corrected and accelerated confidence intervals in [ ]. Confidence intervals and standard errors based on 1000 bootstrap samples

Predictor variable	R^2	$Adj.R^2$	*F*	*p*	*Constant*	*b*	*SE B*	*β*	*p*
Step 1	.388	.384	100.24	<.001	-21.99 [-30.11, -14.03]		4.05		.001
Age in months						1.34 [1.09, 1.58]	0.13	.62	.001
Step 2	.557	.549	65.49	<.001	-31.09 [-40.10, -23.15]		4.34		.001
Age in months						0.66 [0.37, 0.90]	0.15	.31	.001
ESB						0.77 [0.50, 1.13]	0.15	.43	.001
MI						0.28 [0.03, 0.51]	0.13	.13	.019

ΔR^2 = .169 for step 2

As can be seen in Table 24, when only age in months was used as a predictor it accounted for 38.8% of the variance in APS-RVT. The second model, which added ESB, and MI was able to explain 55.7% of the variance (*Adj.* R^2 =.549) and was found to significantly predict outcome, F (3, 156) = 65.49, $p < .001$. As predicted, the ESB added significantly to the amount of change in APS-RVT (b = 0.77 [0.50, 1.13], p = .001) once age had been entered into the model, and MI was also a significant contributor to the model (b = 0.28 [0.03, 0.51], p = .019).

Given that the ESB showed the strongest correlation with the APS-RVT, the regression was re-run entering the predictors hierarchically to examine the amount of variance explained by the ESB. Once age was entered in the first step, ESB explained an additional 16.1% of the variance in the second step. In step 3, MI explained an additional 8%.

A second regression analysis investigated the amount of variance explained by the predictors in the APS-RVT scores for age groups 2 and 3 with SR Lexical included (see Table 25).

Table 25: Hierarchical linear regression analysis of APS-RVT scores for age groups 2 and 3 (n=109), with 95% bias corrected and accelerated confidence intervals in [ ]. Confidence intervals and standard errors based on 1000 bootstrap samples

Predictor variable	R^2	$Adj.R^2$	*F*	*p*	*Constant*	*b*	*SE B*	*β*	*p*
Step 1	.162	.154	20.65	<.001	-17.35 [-32,81 -2.70]		8.11		
Age in months						1.21 [.75, 1.72]	.23	.40	.001
Step 2	.516	.497	25.33	<.001	-19.66 [-33.82, -7.50]		7.60		
Age in months						.10 [-0.34, 0.52]	0.25	.03	.69
ESB						.90 [0.55, 1.44]	0.19	.48	.001
MI						.13 [-0.27, 0.57]	0.19	.05	.440
SR Lexical						.26 [0.15, 0.37]	0.06	.31	.001

ΔR^2 = .354 for step 2

As can be seen in Table 25, age in months explained 16.2% of the variance. The other predictors (ESB, MI, and SR Lexical) entered in the second step explained an additional 35.4% of the variance, which was found to significantly predict outcome F (4, 104) = 27.68, $p < .001$. Only two predictors; ESB (b = 0.90 [0.55, 1.44], p = .001) and SR Lexical (b = 0.26 [0.15, 0.37], p = .001) contributed significantly to the model; while the MI did not contribute significantly. These findings are similar to the results of the regression analysis with ARA-LUI as an outcome measure (section 6.4.2.1). In both cases MI was no longer a significant contributor to the model when the youngest age group was not included in the analysis.

6.4.2.3 Predictors of sentence repetition lexical morpheme score

To investigate predictors of performance on the SR Lexical, the ESB, MI and APS-RVT were used as predictors. Since the SR task is assumed to be informative not only about

children's auditory memory but also children's lexical and morphosyntactic knowledge (Seeff-Gabriel, Chiat & Dodd, 2010) it was expected that children's performance on the APS-RVT would predict SR Lexical score. Based on the mapping theory (Chiat, 2001) and the findings in Chiat and Roy (2008) the ESB was not expected to be a significant contributor to the outcome in SR Lexical. On the other hand, as the MI and SR tests are imitation tasks sharing similar demands in terms of attending to the examiner, turn taking, short term memory and possibly the understanding of others as intentional beings (Dohmen, 2010), it was anticipated that MI would predict SR Lexical score. Table 26 shows the hierarchical linear regression that examined the percentage of variance in SR lexical morpheme score in age groups 2 and 3 accounted for by the predictors.

Table 26: Hierarchical linear regression analysis of SR lexical morpheme score for age groups 2 and 3 (n = 109), with 95% bias corrected and accelerated confidence intervals in [ ]. Confidence intervals and standard errors based on 1000 bootstrap samples

Predictor variable	***R²***	***Adj.R²***	***F***	***p***	***Constant***	***b***	***SE B***	***β***	***p***
Step 1	.127	.119	15.60	<.001	-21.64 [-41.32, -1.44]		10.62		.04
Age in months						1.29 [0.69, 1.93]	0.30	.36	.001
Step 2	.356	.332	14.40	<.001	-13.46 [-35.56, 7.49]		11.42		.237
Age in months						0.45 [-0.19, 1.10]	0.30	.13	.148
ESB						-0.12 [-058, 0.42]	0.24	-.05	.620
MI						0.80 [0.17, 1.20]	0.32	.25	.016
APS-RVT						0.49 [0.20, 0.75]	0.12	.41	.001

ΔR^2 = .229 for step 2

As can be seen in Table 26, age in months explained 12.7% of the variance when entered in the first step. When the other predictors where added to the model (ESB, MI, and APS-RVT), an additional 22.9% of the variance in SR lexical morpheme scores was explained by the predictors (*Adj.* R^2 = .332), which was found to significantly predict outcome, F (4, 104) = 14.40, $p < .001$. In line with expectations, two predictors significantly contributed

to the model: MI (*b* = 0.80 [0.17, 1.20], *p* =.016), and APS-RVT (*b* = 0.49 [0.20, 0.75], *p* =.001). ESB, on the other hand, did not significantly contribute to the model. Since the commonality between MI and SR is the task rather than the content, it was possible that refusal to participate in imitation might be responsible for contribution of MI to SR Lexical. To check this, the analysis was re-run excluding children who refused to participate in the MI task (n = 3). This revealed a different pattern of relations with the APS-RVT being the only significant contributor to the model.

The focus has been on SR Lexical. To check whether performance on grammatical morpheme showed the same relations, the regressional analysis was repeated using the SR Grammatical as the outcome measure. Again, the ESB was not a significant contributor to the model. In addition, the MI also emerged as a non-significant contributor to the model and the APS-RVT was the only significant contributor to the model (*b* = 0.88 [0.39, 1.37], *p* = .001).

6.4.3 Summary

Partial correlations between measures of sociocognition (SCogQ, ESB, and MI) and language measures (ARA-LUI and APS-RVT) for the whole sample controlling for age in months showed highly significant correlations between ESB, MI, and both ARA-LUI and APS-RVT, with correlations ranging from (*r* = .344 to .522) and *p* < .001. The highest correlations were found between the ESB and the two language measures (ARA-LUI, and APS-RVT). In contrast, SCogQ correlated only with the ARA-LUI, *r* = .323, *p* < .001. When SR test scores (SR Lexical, SR Grammatical) were added to correlational analyses for age groups 2 and 3, the ESB showed significant small correlations with both SR Lexical and SR Grammatical (*r* = .260 and .275 respectively), and the MI showed significant moderate correlations with both (*r* = .376 and .310 respectively). SCogQ relations with the two scores of the SR were non-significant.

Regression analyses investigated potential predictors of ARA-LUI, APS-RVT and SR Lexical. Results suggested that children's performance on the sociocognitive measures (ESB and MI) and the SR Lexical were significant predictors of pragmatic language and receptive language. However, motor imitation was only a significant contributor to the models when analysis was performed for the whole sample including the youngest age group (2;0-2;5) suggesting that its predictive value changes during different stages of

development. Interestingly, testing the relative importance of predictor variables showed that the ESB was the best predictor of both language measures (ARA-LUI, APS-RVT). This suggests that the skills assessed in the ESB are important for children's receptive language and expressive use of language.

Investigating predictors of SR Lexical showed that MI and APS-RVT contributed significantly to the model whereas the ESB did not. The APS-RVT was a more important predictor to the model than the MI, suggesting that children's lexical knowledge contributed more to their ability to repeat the sentences than the sociocognitive skills measured in the ESB and MI.

6.5 Low Performing Children

A key purpose of this study was to develop assessments to identify young Saudi children with language delays and the nature of their difficulties. Having established correlational and predictive relations among measures at group level, this section will focus on children at risk according to performance on one or more of the measures to determine (1) pervasiveness and severity (2) whether this relates to parental report of a suspected or diagnosed problem (3) profiles which may throw light on nature of problems and heterogeneity. To check how children performed across measures, z-scores (mean = 0, SD = 1) were calculated from raw scores. As explained in section 5.3.2.3, the Sentence Repetition test was scored using three scores (SR Lexical, SR Grammatical, and TSA). The TSA showed floor effects in age group 2, which suggests that it might not be appropriate to identify children with language problems at this young age. Furthermore, as shown in section 6.4.1 SR lexical and SR grammatical morpheme scores were highly correlated. In addition, reviewing z-scores of SR lexical and SR grammatical morpheme scores of the low performing children revealed very similar performance in both measures which may be due to significant number of refusals among this group of children. Thus, only SR Lexical is reported in this section.

6.5.1 Pervasiveness and severity

6.5.1.1 Children performing low on one task only

Children who scored one or more SD below the mean on one task only are presented in Table 27 and Table 28. Five children were excluded because of missing data on some measures since we could not exclude the possibility that they would have performed low on other tests as well.

Table 27: Numbers of children performing below -1SD on one task according to age group

	Language Measures			**Sociocognitive Measures**			
	ARA-LUI	APS-RVT	SR Lexical	ESB	MI	SCogQ	Total
Age group 1	0	0	n/a	0	2	4	6
Age group 2	1	1	1	1	4	3	11
Age group 3	1	2	3	2	1	1	10
Total	2	3	4	3	7	8	27

n/a: the Sentence Repetition test was not administered to age group 1

As seen in Table 27, a total of 27 children (16.77% of the sample) performed more than 1 SD below the mean only on one task. Examining individual profiles of children in Table 27 showed that for the majority (88.89%, n = 24), parents reported no concern about their children's speech or language development. The three children whose parents reported concerns were low on the language measures. Focusing on performance on language measures showed that for 33.33% of the nine children who performed low on one language measure parents reported a diagnosed or suspected problem. These children may need to be monitored. However, for the remaining 66.67% there might be little reason for concern. Low performance on only one task could be due to a number of reasons other than inability. For example, low scores on direct measures could be attributed to fatigue, or unwillingness. Four children (14.81%) refused to participate in either the MI or SR tasks. A further breakdown of low scores according to severity is shown in Table 28.

Table 28: Numbers of children who performed low only on one task according to z- score

	Language Measures			Sociocognitive Measures			
	ARA-LUI	APS-RVT	SR Lexical	ESB	MI	SCogQ	Total
-1 SD	2	2	1	1	5	3	14
-1.5 SD	0	1	3	1	2	3	10
-2 SD	0	0	0	1	0	2	3
Total	2	3	4	3	7	8	27

As can be seen in Table 28 more than half (55.56%) of the children who performed low only on one language task scored between -1 and -1.5 SD. The remaining children (44.45%) scored between 1.5 SD and 2 SD below the mean and none of these children scored more than 2 SD below the mean. This further shows that children who perform low only on one measure might be those with minor delays who do not require immediate intervention.

With regard to performance on the sociocognitive measures, these measures were selected based on their theoretical and empirical relations with language and communication. Thus, it was expected that low performance on those measures would be more common in children with language or communication problems. However, seven and eight children performed low only on either the MI or SCogQ respectively, with average performance on all language measures. On the MI task most of these children performed between -1 SD and -1.5 SD and as mentioned above some were due to non-compliance. In contrast most of the children who scored low on the SCogQ were more than 1.5 SD below the mean, pointing to the possibility that the SCogQ measure might show high level of false positives.

6.5.1.2 Children performing low on more than one task

Table 29 shows the pattern of performance of children who scored 1SD or more below the mean on more than one task.

Table 29: Children who performed low on more than one task

				Language Measures			Sociocognitive Measures			
Child ID	Age group	Gender	Lang. status	ARA-LUI	APS-RVT	SR Lexical	ESB	MI	SCogQ	No of Tasks
53	1	g	n	✓	✓	n/a	✓	✓	✓	5 (2✓, 3✓)
140	1	g	n		✓	n/a		✓		2✓
149	1	g	n	✓	✓	n/a		✓		3✓
1	1	b	s	✓	✓	n/a	✓	✓	✓	5 (1✓,1✓,3✓)
5	1	b	s		✓	n/a	✓		✓	3 (1✓,2✓)
69	1	b	s	✓	✓	n/a	✓	✓		4 (1✓, 3✓)
89	1	b	s	✓	✓	n/a				2 (1✓,1✓)
43	2	g	n	✓		✓				2✓
122	2	g	n	✓	✓		✓	✓		4✓
128	2	g	m	m			✓	✓	m	2✓
165	2	g	m	✓	✓	✓	✓	✓	✓	6 (3✓, 2✓,1✓)
62	2	b	m	m	✓	✓	✓		m	3 (1✓,2✓)
121	2	b	n	✓		✓		✓		3 (1✓, 2✓)
126	2	b	d	✓	✓	✓	✓	✓		5 (2✓, 3✓)
63	3	g	m				✓	✓		2✓
104	3	g	n		m	✓	✓	✓		3 (1✓,1✓, 1✓)
110	3	g	n			✓		✓	✓	3 (1✓,2✓)
158	3	g	s	✓	✓	✓				3✓
34	3	b	n	✓					✓	2✓
56	3	b	s	✓	✓	✓	✓	✓		5 (2✓,3✓)
74	3	b	n		✓	✓				2✓
94	3	b	n	✓	✓		✓			3 (1✓,1✓,1✓)
97	3	b	s			✓			✓	2✓
135	3	b	s	✓		✓				2 (1✓, 1✓)
157	3	b	n		✓			✓		2✓
Total	25			15	16	12	14	15	7	

n = no concern, s = suspected speech or language delay, d = diagnosed speech or language problem, m = missing (information on language status or test score), ✓ = 1SD or more below the mean, ✓ = 2 SD or more below the mean, ✓ = more than 3 SD below the mean.

As can be seen in Table 29, twenty five children (15.53% of the sample) performed low on more than one task. Among these children, 32% had reported parental concerns about their language development, and just one child (4%) was reported to a have a diagnosis of language delay or disorder. For 48% of these, parents had expressed no concerns about their language development. Information on language status was missing for 16% of these children. Focusing on performance on language measures showed that for 39.13% of the 23 children who performed low on language measures parents reported a diagnosed or suspected problem.

With regard to the gender distribution, the proportion of boys was slightly higher than girls (56% and 44% respectively). Unlike children who performed low only on one measure, children with low performance on more than one task are likely to be at greater risk for continued difficulties.

The most impaired performance across measures was found in children who performed low on *all* tasks, with many of the scores falling more than 2 or 3 SD below the mean, indicating that severity is associated with pervasiveness of difficulties.

Interestingly, there was consistency in performance on the ARA-LUI and the direct language measures with nearly all children who were low on the ARA-LUI emerging low on a direct measure of language. In contrast, performance on the direct and indirect measures of sociocognition was less consistent. As can be seen in Table 29, seven children were low on the SCogQ. Three of these showed severe delays across all tasks. On the other hand, two children performed within average range on ESB and MI.

With regard to compliance on the direct measures, as mentioned in section 6.3.3, a significant number of children were scored 0 on both the Motor Imitation and Sentence Repetition tasks due to non-compliance. Further examination of the performance of those children across tasks showed that for the majority of refusals a pattern of delay was evident on more than one task. More specifically, 77.78% of children who refused to participate on the Motor Imitation test and 66.67% of those who refused to participate in Sentence Repetition scored low on other tasks as well, pointing to the possibility that indeed for some children refusal is an indication of inability or difficulty.

6.5.2 Profiles of low scoring children and their distribution

Sociocognitive measures were included as part of the battery based on the theoretical and empirical evidence of their relations with language. Thus, it is expected that for some children with language delays, difficulties will also be evident in the sociocognitive skills. Examining profiles of low scoring children on more than one task (n = 25) suggested 4 possible profiles:

- Profile 1: Delay only on language measures (ARA-LUI, SR Lexical, APS-RVT) (5 children = 20%).
- Profile 2: Delay only on nonverbal sociocognitive tasks (SCogQ, ESB, MI), (1 child = 4%).
- Profile 3: Mixed pattern of delay (delay on one or more language and one or more nonverbal sociocognitive tasks) (15 children = 60%).
- Profile 4: A pattern of delay across all tasks (3 children = 12%).

1 child showed delay on nonverbal tasks but had missing data for the ARA-LUI (4%).

As seen above, the majority (60%) of low performing children showed a mixed pattern of delay and 12% were low on all measures with many of the scores falling more than 2 or 3 SD below the mean. Only one child was low on the sociocognitive measures only.

Clearly if scores on the SCogQ were not considered – as this measure appeared to be less valid than the other measures - the percentages in the profiles will change.

6.6 Parental Concern Relations to Children's Performance on the Battery of Assessments

Given the significant number of low performing children whose parents did not report concerns, it was of interest to further explore parental report of a suspected or diagnosed speech or language problem in relation to children's performance on the language measures (questionnaires and direct assessments).

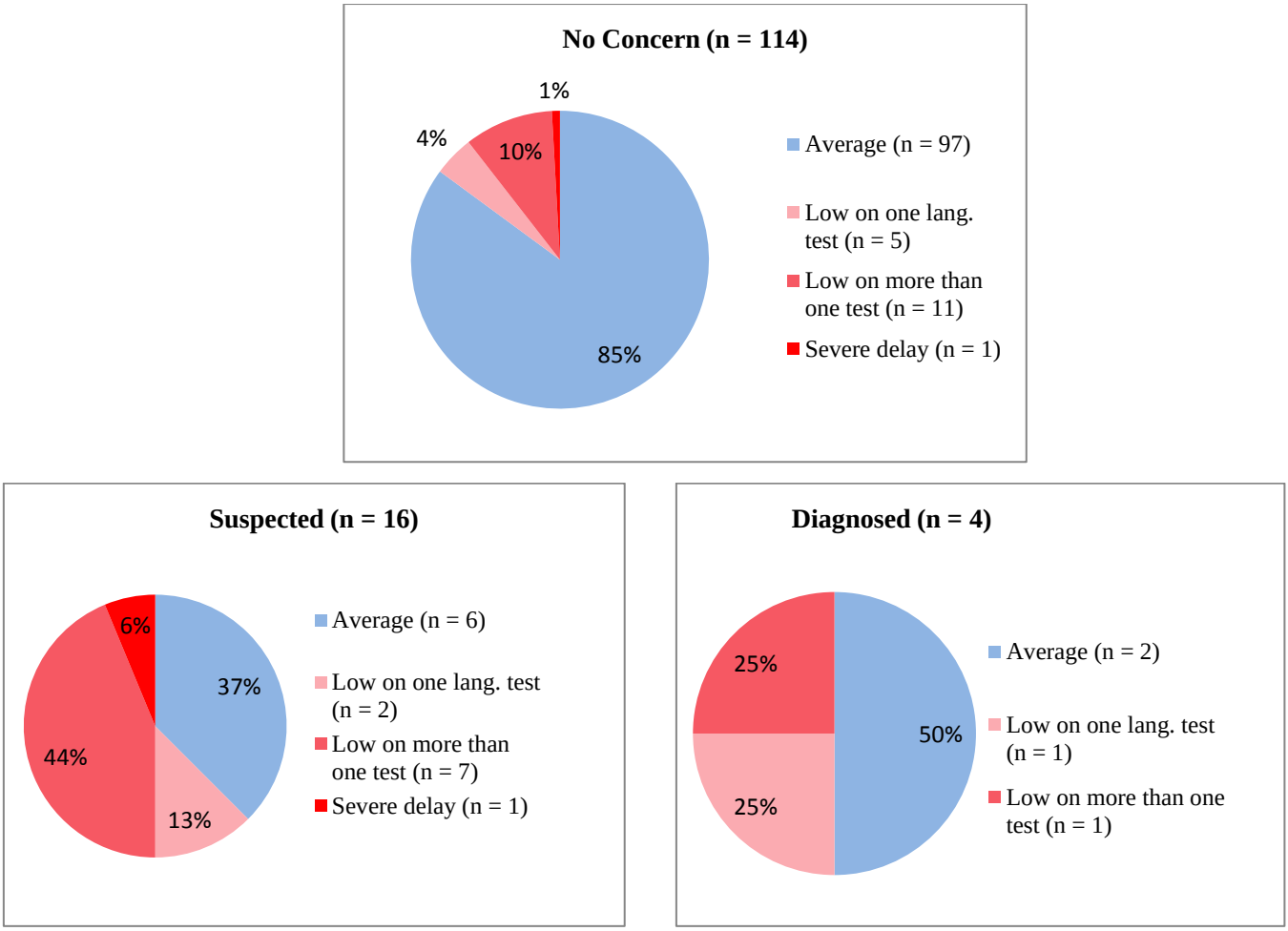


Figure 11: Parental concern in relation to children's performance on the battery of assessments

As can be seen in Figure 11 the majority of parents who reported no concerns (85.09%) had children who performed in the average range on all language measures. Conversely, the majority of parents who reported a suspected speech or language delay (62.5%) had children who performed low on one or more measures of language. In addition, only 4 children were reported to have a diagnosed speech or language problem. Half of these children performed in the average range on all language measures. This may be due to having a speech problem that was not detected by the measures in the battery. Another possibility is that these children may have had an earlier language problem and caught up. Difference between parental report of suspected or diagnosed problem and performance on tests will be further considered in the discussion.

CHAPTER 7

DISCUSSION

The main purpose of this study was to develop a battery of assessments to identify Saudi (Najdi) children with early language delay using carefully selected measures reported to provide best predictions of language and communication outcome and to be informative about children's strengths and difficulties. Due to a lack of systematic assessment tools and normative data in Saudi Arabic, clinicians either resort to informal methods of assessment or translate assessments that were developed and normed for English language and culture without taking into account linguistic and cultural differences. In both cases, clinicians cannot objectively compare the language and communication skills of assessed children to their similar aged peers and decisions made based on these evaluation methods may be inaccurate or misleading. Thus, developing a battery of measures that allows systematic, reliable and informative examination of children's early language and communication skills will greatly contribute to the assessment of young Saudi children. The battery in this project included measures that assess early language and sociocognitive skills. Most of these measures were either systematically adapted or developed taking into account the Saudi (Najdi) language and culture. This process, as described in chapter 4, included drawing new picture stimuli, careful translations taking into account experts' advice, native speakers' views, parents' feedback, and extensive piloting to ensure as far as possible that the assessments are linguistically and culturally appropriate for young Saudi children. Results show that all the measures in the battery are reliable, and most are age-sensitive and therefore show potential for discriminating between typically developing children and children with language or communication delays. In addition, a high compliance rate was found on all measures making them suitable for evaluating language and communication skills of children in the targeted age range.

A second aim of this project was to investigate the extent to which the sociocognitive and language skills related to each other and predicted concurrent language. In the absence of gold standards in Saudi, examining relations between the measures also served to inform concurrent validity. In addition, most previous research on these skills and the way they relate to each other in early language development has been conducted in Western culture with English-speaking children, and the present study is the first to address these issues in

Arabic-speaking children. Furthermore, since this study included a wider range of assessments tapping different sociocognitive and language skills than most previous studies, it was possible to investigate in more detail interrelations among the various developmentally important language and sociocognitive skills than in studies with a more restricted range of measures. Thus, for both theoretical and practical reasons, findings from this study have the potential to provide more insight into the different skills that underpin language and may give rise to language and communication problems.

The first section of this chapter will summarize and discuss the main findings of the study. Children's performance in relation to parental concern is discussed in the second section. The third section focuses on discussing the role of sociocognitive skills in language. In the fourth section, methodological challenges and limitations are discussed, followed by clinical implications in the fifth section. Finally, the last section explores directions for future research.

7.1 Main Findings

7.1.1 Reliability and validity

Excellent levels of test-retest and inter-rater reliability were found for the parental reports and direct measures. Caution is advised however in interpreting reliability levels in parental reports as Fenson (1994) noted that artificially high levels of test-retest reliability may arise due to parents consistently over-estimating their child's abilities or simply remembering their previous answers. In order to establish reliability for parental reports, Fenson highlighted the importance of demonstrating their concurrent validity with direct and independent methods of assessment.

Due to the lack of previously validated measures tapping the same skills in Saudi Arabic, concurrent validity was measured by investigating how the measures used in this project correlated with each other. Correlations between the parental reports of language and direct measures of language (ARA-LUI, SR Lexical, SR Grammatical, and APS-RVT) partialling out age in months showed that the majority of measures were significantly and moderately correlated despite the fact that the measures differed in the method of assessment, aspect of language being tested and procedure of administration. These

findings provided sufficient evidence on the validity of the language measures and suggested that they documented interrelating language skills. Likewise, correlations between the direct and indirect measures of sociocognition (SCogQ, ESB, and MI) partialling out age in months showed significant moderate correlations between the ESB and Motor Imitation. Surprisingly, however, non-significant correlations were found between the SCogQ and both the ESB and the Motor Imitation which may indicate that parents may be less able to judge children's social communicative behaviors (in the SCogQ) than their use of language (in the ARA-LUI). See further discussion in section 7.4.3.

Further evidence for validity was the finding that all the measures (with the exception of the SCogQ) were age sensitive, and the scores of the children in the current study were in agreement with previous research (Chiat & Roy, 2006; Dohmen, Chiat & Roy, 2013; O'Neill, 2007). See further discussion in section 7.4.2.

To sum up, the reliability and validity results suggested that all but one of the novel and adapted tools generated psychometrically sound measures that were fit for purpose, the possible exception being the SCogQ.

7.1.2 Relations between sociocognitive skills and language

A key motivation for this study was the evidence, in line with the sociocognitive hypothesis, that deficits in early sociocognitive skills are specifically predictive of difficulties with pragmatic language and social communication (Chiat & Roy, 2008, 2013). The combination of measures employed in this study was selected to build on this evidence by addressing two key questions:

- Do the sociocognitive skills assessed in the ESB (social responsiveness, joint attention and symbolic comprehension) uniquely predict concurrent performance on different language outcomes (pragmatic language, receptive vocabulary, sentence repetition)?
- Does the sociocognitive skill of motor imitation show unique predictive relations to these different language outcomes, making a contribution additional to or distinct from the ESB?

In order to address these questions, regression analyses were conducted entering the two measures of sociocognition as predictors simultaneously with receptive vocabulary and/ or sentence repetition lexical morpheme score to see if the sociocognitive skills would contribute to language outcome even when other aspects of language had been taken into account.

Important findings emerged from these analyses. First, the ESB was the strongest predictor of pragmatic language and receptive vocabulary when the analysis was performed on the whole sample as well as when it focused only on age groups 2 and 3. MI was also a significant contributor to the model when the analysis was performed using the whole sample, although its additional predictive ability once ESB had been taken into account was small. However, MI was no longer a significant contributor to the model in the analysis that excluded the youngest age group. These findings echo the vast amount of research in Western cultures reporting relations between sociocognitive skills and language in typically developing children (Beuker, Rommelse, Donders, & Buitelaar, 2013; Doswell, Lewis, Boucher & Sylva, 1994; O'Reilly, Painter, & Bornstein, 1997; Tomasello & Barton, 1994; Tomasello & Farrar, 1986; Tomasello, Strosberg, & Akhtar, 1996), late talkers (Thal & Tobias, 1992; Thal, Tobias & Morrison, 1991) and children with autism (Thurm, Lord, Lee, & Newschaffer, 2007; Toth, Munson, Meltzoff, & Dawson, 2006), although most of the research focused on younger children. Thus, the current results suggest that these early sociocognitive skills continue to influence language outcome up to 3;5 years of age.

A different picture of relations between sociocognitive skills and language (in the two oldest age groups) emerged when sentence repetition lexical morpheme score (SR Lexical) was used as the outcome measure. The ESB did not contribute to outcome in SR Lexical. MI showed unique relations with SR Lexical, although the greatest impact on the model was made by the APS-RVT. Similar findings were reported by Chiat and Roy (2008) in their longitudinal study with clinically referred children: the ESB did not add to the variance explained in sentence repetition when entered with other measures of phonology, receptive and expressive language. Implications of these findings for our understanding of the role of the sociocognitive skills to language will be discussed in section 7.3.

7.2 Parental Concern in Relation to Children's Scores on the Different Measures

In Saudi, referrals to speech-language therapy clinics for children with speech-language delays with no other obvious congenital, genetic or neurological problems are likely to come from parental concern as there is no system for identifying and referring these children through the medical or educational settings. Thus, an important factor in deciding whether a child is enrolled in early intervention services or not is parents' awareness of speech and language impairments. This section will focus on parental concern in relation to children's performance on the language measures.

In this study parents were asked at the end of the ARA-LUI whether or not they suspected that their child had speech or language delay and whether or not their child had been assessed by a professional. Relations between parents' report of suspected or diagnosed problems and children's performance on the language measures (questionnaires and direct assessments) were analyzed (see section 6.6 and Appendix K). Results showed that while the majority of children (85.09%, 97 out of 114) whose parents reported no concerns about their speech and language performed in the average range on all language measures, 14.91% (17 out of 114) performed low on one or more measures of language including a child who showed broad delays spanning all direct and indirect measures of language and sociocognition. There are a number of possible reasons for these findings. One is that parents might have been reluctant to disclose such information in a research context. Informal observations showed that some parents of low performing children who reported no concern did approach the investigator either directly or through the child's teacher and asked for feedback on their children's performance which might suggest that they were concerned about their children's language development. Interestingly, two more parents of children who showed delays in both sociocognitive and language measures left the question about concern unanswered which may further indicate that some parents were conservative in answering this question. Alternatively, parents may not be aware how their children compare with their typically developing peers and did not realize that their children might be delayed. Another possibility is that these cases represented true cases of 'no reason for concern'. It is important to note that about one third of these children performed low on only one measure of language. Acknowledging that performance in this young age is highly influenced by motivation, compliance and attention points to the possibility that some of these children are indeed cases of true no concern.

Conversely, the majority of the children whose parents reported a suspected speech or language delay (62.5%, 10 out of 16) performed low on one or more measures of language, although over a third of these children performed within average limits on all language measures. This mismatch is more than twice the mismatch for 'no concern' children. It is difficult to determine whether these children showed problems that were not detected by the measures in the battery, for example, speech difficulties or problems with other aspects of language, or whether they were cases of over-concerned parents. In interpreting the findings of rate of parental concern it is important to remember that the question about concern was addressed in the context of completing a questionnaire about children's language (ARA-LUI) which, as noted by Klee, Pearce and Carson (2000), may increase parents' awareness and prompt their expression of concern. In addition, the present findings may have been affected by participants' SES. The sample in this project was biased towards middle to high maternal education. Keegstra, Knijff, Post and Goorhuis-Brouwer (2007) suggested that parents with a high level of education tend to be over-concerned about their children's language.

With regard to parental report of a diagnosed problem, two out of the four children in this group performed in the average range on all measures. Again, these children might have speech only problems or might have shown earlier delays and caught up. In fact the parent of one of these children reported that the diagnosis was made more than one year before the child participated in the current study, so it is possible that this child was a "*late bloomer*" (Rescorla, 1989).

In sum, it appears that a substantial number of parents of low performing children had concerns about their children's language development. The rate of agreement between parents' concern and language scores in the current study is very similar to that reported for 3 year old children with early expressive language in Bishop, Price, Dale and Plomin's (2003) study. Using parental reports of language to identify language impairment, the authors found that 64.1 % of children whose parents reported concern performed low on 2 out of 3 measures of language. However, in Bishop and colleagues' study there was a larger mismatch between parents who reported no concern and children's scores, with 32.6% of children in this group performing low on language measures.

Parental concern is not only important in identifying children with language delays, but also has been suggested to improve the predictability of language outcome one year later

(Klee, Pearse & Carson, 2000). However, expressing concern about language delay does not guarantee that parents will seek professional help. In fact, while 9.94% (16 out of 161) of the parents had reported that they suspected that their child had a speech or language problem in the current study, none of those had contacted a speech language therapist. Factors such as parents' awareness of long term effects of speech-language problems and the importance of early intervention, perception of speech-language problems, level of concern, tolerance of such problems, and accessibility of consultation services will have an impact on the number of children seeking intervention. Bishop and colleagues (2003) reported that although there was an overlap in their study between the children whose parents reported concerns at 3 and 4 years of age and those who sought professional consultation, the overlap was not complete, with 33% of 3-year-olds and 13% of 4-year-olds with parental concerns reporting no contact with a professional which may further suggest that the age of children might have an effect on whether the concerned parent seeks professional advice or not. The overlap between parental concern and contacting a professional in our sample was very small with the majority of parents (80%) reporting no contact with a professional.

While we do not expect a perfect match, it is evident from the current findings that based on parental concern many children with speech language problems were not identified early and enrolled in intervention services. Relative to the whole sample with complete data (n = 134), 12% of the children whose parents did not express concern showed delays on the language measures of the study compared to 4.5% of children whose parents reported a suspected problem performing in the average range. Thus, it appears that more children might be missed in comparison to over referrals. Given that the sample came mainly from middle to high SES, the present findings may indicate that parents in Saudi may be more tolerant of early language delay and are less aware of importance of early intervention (see further discussion in section 7.4.2). This calls attention to the importance of improving parental awareness and referral services to speech-language clinics in Saudi. A first step in improving referral services might be to work with nurseries to educate teachers about red flags for language and communication problems.

7.3 The Role of Sociocognitive Skills in Language

A second aim of this study was to examine the associations and dissociations between sociocognitive and language measures in a large group of Saudi children to improve our understanding of language and communication development. Important findings emerged from the regressional analyses that looked at these relations using a wider range of measures than most previous research. Implications of these findings for our understanding of the role of sociocognitive skills in language are discussed in this section.

7.3.1 Importance of sociocognitive skills for language development cross-culturally

First, results of the current study suggest that language difficulties may be related to sociocognitive difficulties. Regressional analyses showed strong unique relations between sociocognitive skills and different language outcomes even when other aspects of language had been taken into account (see sections 6.4.2 and 7.1.2). These findings corroborate previous research with late talkers (Desmarais et al., 2008; Thal et al., 1991; Thal & Tobias, 1992) suggesting that sociocognitive skills may be linked to language skills in important ways and thus are important potential predictors of language outcome. The strong relations were observed in a very different language culture than most previous studies, and one which may have shaped different parental practices and in turn developmental outcomes (Lin & Chiu, 2014). Hence the study provides new evidence on the importance of sociocognitive skills for language outcome transcending cultural and language differences. The fact that these relations were also found when different methods for assessing language were used (i.e., direct assessment and parental reports) adds to the value of these findings.

Furthermore, the strong links between sociocognitive skills and language observed in the current study echo the vast amount of research of children with autism. Thus, in accord with Luyster and colleagues' (2008) proposal, this may suggest that the process of language development is similar in both typically developing children and children with autism in terms of its underlying framework.

More specifically, and in line with the sociocognitive hypothesis, the findings suggest that sociocognitive skills are specifically informative about social communication abilities. The ESB made the strongest impact on outcome in pragmatic language and receptive vocabulary when other aspects of language had been taken into account. On the other hand, the ESB was not a unique predictor of outcome in SR Lexical. Interestingly, these

findings are strikingly similar to findings of Chiat and Roy (2008) with their clinically referred children despite the fact that different measures for assessing language were used in the two studies. Chiat and Roy also reported that the ESB emerged as a significant predictor of outcome in pragmatic language when other measures of phonology, receptive and expressive language were taken into account. In addition, the ESB in Chiat and Roy's study did not add to the variance explained in sentence repetition when entered with other measures of language. The only difference in findings was that the ESB was not a significant predictor of later outcome in receptive language in Chiat and Roy's study when entered with other measures of receptive and expressive language. This may reflect several differences between the studies. First, Chiat and Roy's study was a longitudinal study. Second, the same measure of receptive language was used among the predictor variables and as an outcome measure when children were first seen and at follow-up which could have overshadowed what the ESB was adding to the model. It is also important to remember that the children in Chiat and Roy's study were clinically referred with concerns about language. However, based on performance of the low performing children (section 6.5) indicators so far are that a large proportion of the children with a low score on the ESB showed impaired performance on receptive vocabulary as well as pragmatic language.

Relations between the sociocognitive skills and later outcome were not investigated in the current study, however, the ESB assessed skills considered as precursors to the ability to attribute mental states to others (i.e. theory of mind) (Tomasello, 1995). Based on the notion that theory of mind (ToM) forms the infrastructure of several pragmatic abilities such as referential communication, understanding and use of irony, conversational repair and use of modals and mental state terms (Dahlgren & Sandberg, 2008; Martin & McDonald, 2004; O'Neill & Atance, 2000), we expect that children's performance on the ESB will relate to outcomes in pragmatic language later in life. Further support for the long term predictive value of the ESB is provided by the findings reported in Chiat and Roy (2008, 2013) that impairments in sociocognitive skills assessed in the ESB predicted social communication at 9-11 years of age. Results also suggest that deficits in the skills assessed in the ESB might be related to communication difficulties that do not necessarily qualify as ASD.

7.3.2 Contribution of motor imitation and changes with age

Turning to MI, findings suggest that MI is independently informative about language outcome. MI emerged as a unique predictor of outcome in pragmatic language and receptive vocabulary when the analysis included the whole sample. In addition, unlike the ESB, MI was a unique predictor of outcome in SR Lexical in the two oldest age groups (see sections 6.4.2 and 7.1.2). These findings provide evidence that MI is picking up on additional skills beyond those assessed in the ESB. As mentioned in chapter 3, relations between MI and language have been attributed to different underlying skills such as the understanding of others as intentional agents (Carpenter et al., 1998; Meltzoff, 2011) and representational requirements (Stone et al., 1997). However, if we assume that those skills are measured in the ESB, further explanation of the distinct role of imitation in language is warranted. One possibility is that MI is additionally picking up on the child's motivation to take an active role in interpersonal exchanges which might be considered important for creating the framework for pragmatic development. Children who show this behavioral profile will have longer interactions with a communicative partner. Thus, they have more opportunities to map forms to meanings (Chiat, 2001) in contexts were joint attention has already been established. Accordingly, these children will also have better receptive vocabulary. That might also account for the stronger impact of MI on SR Lexical when compared to the ESB, as both measures (MI and SR) require the child to be motivated to interact with the examiner. This explanation is in line with Uzgiris's (1981) view of imitation as a reflection of interest in interpersonal interaction. To support this role of MI is the fact that poor performance on MI did not appear to be due to inability to match the model correctly but rather due to refusing to imitate. Refusal in this case seems to be an indication of the child's lack of desire to engage with others. It follows from this position that children who refused to participate in the MI task will also perform poorly on the ESB since the latter is considered an informative measure of social engagement. Indeed when correlations between measures were examined (see section 6.4.1), the highest correlation for the MI task was with the ESB. Likewise, Dohmen, Chiat and Roy (2013) argued that selective refusal among the specific language delay (SLD) group in their study in imitating postures and gestures as opposed to the imitation of instrumental tasks might be an indication of specific difficulty with the task rather than uncooperativeness.

In line with this argument regarding the role of MI are the findings of Desmarais and colleagues' (2010) of relations between engagement in communication and receptive and

expressive language abilities in their late talkers group. In addition, the current findings on the role of MI corroborate a recent small study of children with autism (Miniscalco, Rudling, Råstam, Gillberg, & Johnels, 2014). Based on parental reports the authors found that both children's vocabulary and imitation of adult actions correlated with their pragmatic language. However, only imitation of adult actions emerged as a significant predictor of pragmatic language. In contrast, Dohmen (2010) found no correlations between motor imitation and receptive language in the two youngest age groups (2;0-2;5, 2;6-2;11) in her clinical sample and only in the oldest age group (3;0-3;5) were there significant correlations. This could be because relations between receptive language and motor imitation in Dohmen's study were explored only for clinically referred children who, as the author argued, may have shown limited variations in receptive language abilities especially in the youngest age groups; only in her oldest age group did receptive language appear to be more variable.

Taking a closer look at relations between MI and language across age groups in the current study suggests that MI plays a transitional role in language development. In contrast to the ESB, MI was no longer a unique predictor of pragmatic language or receptive vocabulary when the analyses focused only on the two oldest age groups (see section 6.4.2). Early in development motor imitation might enable children to have long lasting exchanges with a communicative partner before the emergence of words. Thus, imitation appears to provide a means towards developing referential communication. However, this role is transitory (Nadel, Guérini, Pezé & Rivet, 1999). Once children have mastered this skill other sociocognitive skills are more crucial for the continued development of language (Toth et al., 2006). On this view, the role that motor imitation plays in language development appears to be similar to that of communicative gestures. Gestures provide children with a tool to communicate, therefore facilitating verbal development (Goodwyn, Acredolo, & Brown, 2000). Gestures may also provide a window into the processes that underlie word learning (Goodwyn, Acredolo, & Brown, 2000). Examining gestural abilities of late talkers may throw light on their language difficulties and predict risk for long term problems. In accordance with this argument, late talkers who caught up were reported to use more gesture to compensate for their limited verbal language than late talkers with persistent language problems. Interestingly, and in line with our findings on relations between MI and receptive language, late talkers who showed poor gestural abilities whether in an imitation or production task were impaired in receptive language skills as

well (O'Neill & Chiat, 2015; Thal et al., 1991; Thal & Tobias, 1992). In this regard, it would be interesting for future research to examine if gestures, like MI, would be less predictive of language at a later age compared to other sociocognitive skills.

Given the diminishing role of MI with age, why was it still a unique predictor of SR Lexical in the two oldest groups? It was argued earlier that this relation might be due to the fact that both MI and SR are direct measures that depend on the child's engagement with the examiner. Since both are imitation tasks, they may have more in common. Decety (2006) argued that imitation is not a simple matching behavior but rather a molar construct that includes different subcomponents such as perception-action coupling, visual attention, short-term memory, body schema, mental state attribution and agency. Thus, the relation between MI and SR Lexical performance might be due to the fact that both tasks require attention, attribution of mental states, perception of stimuli (auditory or visual), the ability to hold the stimuli in short term memory, turn taking and the planning and execution of the perceived stimuli via a motor act. Running the analysis again excluding the children who refused to participate in the MI task (n = 3) helped clarify the nature of associations between MI and SR Lexical. Taking out the refusers, who are hypothesized to have difficulties with interpersonal engagement, MI ceased to contribute significantly; APS-RVT was the only significant predictor. This suggests that once children are willing to imitate, their performance on sentence repetition (as measured by SR lexical) is largely determined by linguistic abilities.

7.3.3 Summary

To summarize, the current study suggests that the ESB is a very important predictor of language outcome. The contribution of the ESB was larger than the contribution of other measures of language. It seems that the range of skills assessed on the ESB – from social responsiveness and joint attention which are necessary to understand others' intentions and therefore a prerequisite for understanding the meanings behind their words, to the symbolic comprehension task which draws on the capacity for understanding symbol-referent relations – uniquely influence pragmatic language and receptive vocabulary across the full age range. Motor imitation is also an important predictor of language development. MI seemed to be tapping active interpersonal engagement. Accordingly, MI can be considered as a foundation for communication, but once children attain this, other sociocognitive and language-processing skills become crucial for language development.

7.4 Issues to Be Considered in the Interpretation of the Results

7.4.1 Inclusion criteria

The recruited sample included children with no concerns about language and communication development as well as children with concerns about language development and those who were reported to have a diagnosed language problem. Thus, the sample might be more representative of the general population and not of a typically developing sample. Furthermore, since this is a self-selected sample, parents with children with concerns about language development might have been more interested in participating than parents of typically developing children. Hence, the sample might have included more low performing children as compared to the general population (see further discussion in the next section). But given the lack of information on parents who did not agree to participate, this cannot be confirmed.

Despite possible biases in the sample, results indicate that the sample represented children with a broad range of abilities which is very useful as a first step in exploring the potential of the newly developed or adapted tests. Pena, Spaulding and Planet (2006) have argued that, although including children with language impairment might be useful in showing how a child performs in reference to a general population, it reduces the diagnostic accuracy of a test. The authors conducted a review of test manuals and a simulation study that compared normative samples including and excluding clinical cases. They found that inclusion of clinical cases resulted in more overlap in performance between clinical children and the normative sample, a smaller mean group difference, and larger standard deviations and variability in performance, lowering the test's sensitivity. However, this argument was based on reviews of language tests which for the most part were for children older than 3 years of age and their simulation study was based on prevalence of LI in 5 and 6 years old children in which less variability in typical children's language is expected. Children in the current study ranged in age from 2;0-3;5 years. Accordingly, to develop norms for each measure, future research might consider whether or not children who performed below the 3rd percentile should be excluded. In children over 3 years of age, less variability might be expected based on evidence of the significant number of late talkers who caught up by age 3 (Dale et al., 2003) and studies that reported a vocabulary spurt in late talkers occurring at different points between the ages of 2;0 and 3;0 years of age (Rescorla, Mirak, & Singh, 2000). On the other hand, for children below 3 years of

age more variability is expected, so excluding these children may result in skewing up the data and over-identifying children who might be performing on the lower end of a normal range.

It is also worth noting that children's nonverbal abilities were not considered in the exclusion criteria. While this is in line with the inclusion criteria in the norming studies for parental reports such as the CDI (Fenson et al., 1994) and the LUI (O'Neill, 2009), children's nonverbal abilities have sometimes been considered in norming studies for direct language measures to ensure that the norming sample represents the population in terms of distribution of children with intellectual impairments. IQ is also usually among the exclusion criteria of small scale late talkers' studies, although different cut-offs for nonverbal IQ have been used in different studies. Interestingly, researchers have recently been arguing that children's performance on IQ tests may not be considered essential in identifying children with specific language delays. For example, Dockrell and Marshall (2015) used results of studies that reported similar responses to oral language intervention for children with and without discrepancies between their verbal and nonverbal abilities as evidence against the necessity of characterizing children based on their nonverbal abilities. In addition, Gallinat and Spaulding (2014) pointed out that it is difficult to rule out that children's performance on nonverbal IQ tests is biased and affected by their linguistic skills as children's poor language may influence their ability to follow the test's instructions as well as their reasoning abilities. Furthermore, there is evidence that nonverbal IQ is not stable in language impaired children and may change with time (Botting, 2005; Chiat & Roy, 2008). However, Bishop (2014) noted that nonverbal abilities may be important to consider when the purpose of the research is to identify correlates of language deficits, to be able to demonstrate that the observed relations are not modulated by the cognitive abilities. Accordingly, findings of this study must be considered in light of potential limitations.

7.4.2 Age sensitivity

Variation in performance and discrimination according to age are vital for diagnostic assessment. Children's performance on five of the six measures of the battery (ARA-LUI, ESB, MI, SR, and APS-RVT) was similar in the following aspects:

1. Descriptive statistics showed wide distributions in the performance of children belonging to the same age group.
2. An increase in mean scores of children from the three age groups was observed.
3. One-way ANOVA analyses further confirmed that the measures were age sensitive.

In contrast to other measures, the SCogQ did not show an age effect: mean scores of children in the three age groups were very similar. Given that the items included in the SCogQ assess skills that for the most part have been reported to appear before the 2nd birthday (Carpenter et al., 1998; Watt, Wetherby & Shumway, 2006) this finding was expected, especially when the items were scored on a 3-point Likert scale which might not be sensitive to differences in the frequency of children's use of these skills. In contrast, the ESB and MI were administered according to a fixed testing protocol and scored not only for presence/absence of a skill, but also the frequency with which this skill was used. They therefore detected how children responded to a given number of stimuli that assess a given skill. The finding that both the ESB and MI yielded differences between the three age groups suggests that, although most of the skills being assessed in both tasks are observed in children in the sample as young as 2 years old, their frequency increased with age. Dohmen (2010) also reported significant differences on posture imitation between children in the age groups 2;0-2;5 and 2;6-2;11 and on gesture imitation between age groups 2;6-2;11 and 3;0-3;5 in both her typical and clinical samples. Further analysis of the subtasks of the ESB showed that the age effect was not observed on all tasks: on the social responsiveness task, no age effect was found, which is consistent with the findings of Chiat and Roy (2006a). This suggests that among the three subtasks of the ESB, social responsiveness was the most basic interpersonal skill and children who were delayed on this task might be those with the most serious delays. Indeed, the children with very low scores who appeared as outliers on the social responsiveness task were those who performed low on all other tasks.

In considering the wide range of scores and overlap observed in the three age groups on the five measures, and making comparisons with other studies, it is important to remember

that the sample included children with concerns about language development as well as children with diagnosed speech or language problems. The reasons for including these children were discussed in section 5.1.2. However, this clearly affects the distribution of the sample: if the study had included only typically developing children, a narrower range of scores would be expected on all tasks. Further inspection of scores, especially on the language measures, points to the possibility that the inclusion of children with concerns about language development resulted not only in a wide distribution but also a large overlap of scores of children from different age groups. For example, examining the scores on the APS-RVT showed that the minimum score of children in age groups 1 (2;0-2;5) and 2 (2;6-2;11) was 0 and in both cases children who scored 0 did not appear as outliers. Furthermore, the range of scores on the ARA-LUI for children in age group 2 was wider than age group 1 with scores in the lowest range in age group 2 being lower than scores of children in age group 1. It is possible that the sample included more language delayed children in age group 2 than age group 1.

Finally, comparing the present results to results of previous studies revealed a similar trend of growth with age for the ARA-LUI, ESB, SR, although the range of scores differed slightly between the present study and previous studies. For example, comparing the median score of children from the three age groups on the ARA-LUI in the current study to the score at the 50th percentile for boys and girls in the corresponding age range on the LUI (O'Neill, 2009) showed that there was a trend for scores of children in the current study to be lower than the Canadian sample on the original LUI (see Appendix J). The LUI norming study also included children with suspected speech or language problem or delay but children with a diagnosed language problem were excluded. However, inspection of performance in the current study revealed that some children who performed at a very low level were not reported to have a diagnosed language problem and were not necessarily among those with suspected language problems. It is possible that the level of awareness of language and communication impairments is lower among Saudi than Canadian parents and that Saudis may be more tolerant of these problems in young children. Likewise, mean scores were slightly lower for SR as compared to Wallan, Chiat and Roy's (2011) study which only included typically developing children with no concern about language development (see Appendix J). Comparing the differences in scores between the current study and previous studies for both SR and ARA-LUI showed that in age group 3 scores in the current study were 10% lower than scores of previous

studies, while in age group 2 scores were lower by 13% for ARA-LUI and 6-9% for the SR Lexical and SR Grammatical scores. The consistent direction and similar magnitude of differences for the two measures further indicate that our sample might have included more low performing children.

In sum, despite the few observed differences, the present data map onto findings from previous studies quite well which supports their validity. In addition, the finding of differences between the three age groups for the measures ARA-LUI, ESB, MI, SR, and APS-RVT shows their potential for identifying children with language or communication delays.

7.4.3 Return rate on parental questionnaires and accuracy of responses

Parental questionnaires provide the opportunity to evaluate skills that are difficult to observe in a direct assessment and to collect information about children's performance in a variety of contexts (Dale, 1996). In English parental reports are commonly used as tools for screening or assessment of young children's language or communication skills (Dale, 1996), but such tools are still lacking in Saudi Arabia. The current study included two parental questionnaires, the ARA-LUI and the SCogQ. This section will focus on addressing the following issues: viability in terms of parental compliance, and reliability in terms of consistency with results on direct measures. Other factors which may influence accuracy such as who fills in the questionnaires will also be discussed.

The two questionnaires (ARA-LUI and SCogQ) varied greatly in length (ARA-LUI: 180 items, SCogQ: 18 items) but since they were sent together to parents, the majority of parents either returned both or neither, so most missing data was for both questionnaires. With regard to the ARA-LUI, 26 questionnaires were not returned (16.77% of the sample), while 25 SCogQ were not returned (15.53% of the sample). Thus, one of the difficulties faced during this project was in collecting questionnaires to reach the targeted number of participants. The rate of return and likelihood of parents completing the questionnaires was anticipated from the pilot study and from nursery supervisors who expressed their skepticism about getting a good return rate and shared their negative experience of parents when filling in forms or questionnaires. However, as already discussed in section 6.1.2.1, parents in this case were filling out the questionnaires for research purposes, therefore they

have a very different motivation from parents completing a questionnaire in a clinical diagnostic setting. Based on personal experience, parents in Saudi usually fill out forms or questionnaires such as case history questionnaires in a clinical setting willingly, so the above mentioned problems may be specific to the use of questionnaires for research purposes. With respect to the ARA-LUI, in some cases parents expressed concerns about the length of the questionnaire. Again, this concern cannot be generalized to parents of clinically referred children.

In order to maximize accuracy of parents' responses, contact information was provided to parents if they needed to inquire about items in the questionnaires. However, very few parents approached the investigator. In a few cases parents completed the questionnaire while attending the direct assessment session. In these cases it was helpful to clarify items that parents found ambiguous or unclear, though such items were few and differed between parents. Future research may further explore whether particular items are problematic.

Consistency of responses between items in the questionnaires that are similar to those assessed in the direct assessment was not systematically analyzed. However, qualitative observations revealed very few cases of mismatch between parents' judgment and children's performance during the assessment session. Thus, it can in general be assumed that parents' responses were reasonably accurate. In addition, as mentioned in section 7.1.1, the ARA-LUI at a group level showed significant moderate concurrent correlations with direct language measures indicating that parents were able to report on their children's use of language with reasonable accuracy. Based on these findings, and the sensitivity of the ARA-LUI to age (see sections 6.3.1.1 and 7.4.2), it may be concluded that it is a valid assessment.

On the other hand, the SCogQ, which was designed to yield information on children's communicative and social skills, showed non-significant correlations with the direct measures of sociocognition (ESB and MI). High correlations between the three assessments were not expected given the differences in the content of the measures, the contexts in which the skills were observed, and perspectives of the parents as opposed to the researcher. Nonetheless, it was expected that low and high performing children would be identified similarly by the three measures, so that some association between the measures was anticipated. There are a number of possible explanations for why this was

not the case. First parents might have misinterpreted some of the items in the questionnaire. For example, when asked "does your child respond to his/ her name" they may have responded based on whether or not the child responds *verbally* to his/her name. Another possibility is that some of the typically developing children in the oldest group received a lower score on some of the items that they no longer perform spontaneously most of the time. For example, when children start using words they may rely less on pointing, affecting the test's ability to discriminate between typically developing children and those who might be at risk.

It is also possible that in comparison to completing the ARA-LUI parents were less accurate in judging their children's social and communicative skills. Although both questionnaires used mainly a recognition-based format, as pointed out by Stiles (1994), the decision strategies employed by the parents differ depending on the skill being assessed. While parents on the ARA-LUI were required to report on their children's language use which mainly involved recalling if a child produces a certain word or sentence (apart from the use of gestures section which is not scored), on the SCogQ parents had to report on social and communicative skills such as gaze-following, point-following and imitation. Judgment of these skills might be more difficult for the parents. For example, a child who follows his/her parent's gaze/point might be responding to an accompanying verbal prompt "look" and not necessarily sensitive to his/her parent's gaze direction and it might be difficult for the parents to discriminate between the two scenarios. Previous studies with similar parent report tools such as the CSBS-DP have shown concurrent moderate to large correlations between parent reports and direct methods of assessment for children 12-24 months of age (Wetherby, Allen, Cleary, Kublin, & Goldstein, 2002). Nevertheless, analysis of the relations between the different domains showed that correlations between the speech composites was stronger than between the social composites leading the authors to suggest that parents may be more able to report on use of words in comparison to social communication. A similar suggestion was made by Eadie and colleagues (2010). In this study with 12-month-old infants, the researchers investigated the validity of the CSBS-DP. Based on the finding of smaller correlations between scores on the parent reports and direct assessments in comparison to those reported in Wetherby et al.'s (2002) validation study, the authors suggested that it may be easier for parents to report on their children's more "overt behaviors" such as the use of words, which are more frequent in older children, in comparison to prelinguistic skills such as gesture use, emotion and eye

gaze, which are the dominant communicative behaviors at 12 months of age. Likewise, the poor sensitivity (35.1%) of the CHAT (Baird et al., 2000), a screening instrument for autism assessing similar aspects to those in the SCogQ such as joint attention and pretend play, suggests that it may be difficult for parents to judge sociocognitive skills. Thus, future research should investigate the possibility of the questionnaire being more effective for the youngest age group, children at risk or in an interview format.

Finally, it is important to note that accuracy of responses for both questionnaires might have been affected by who completes the questionnaire: issues such as the amount of time the informant spends with the child and their SES are important to consider. In this study almost all questionnaires were completed by mothers and only one questionnaire was completed by the child's father. However, it is not known how much time the child spent with the mother and whether the child was cared for primarily by a nanny as is not uncommon for families in Saudi. Furthermore, the possible impact of SES had on response accuracy was not investigated. There have been some suggestions that parents' reporting accuracy of their children's language abilities might be related to SES, but results have been inconsistent. Despite concerns that parents of lower education might overestimate their children's abilities (Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky, & Paradise, 2000), this was not the conclusion reached by Sachse and Von Suchodoletz (2008) using a German version of the CDI. Sachse and Von Suchodoletz reported similar relations between parents' reports and direct methods of assessment for children from different educational backgrounds. Likewise, Pan, Rowe, Spier and Tamis-LeMonda (2004) found no association between maternal education and scores on either the CDI or standardized tests and suggested that reporting accuracy was similar for parents from different educational backgrounds in their low income sample. Thus, this matter is worth investigating in a future study to determine if the SES of parents in Saudi has an effect on their response accuracy.

7.4.4 Children's compliance

Compliance is usually a concern when assessing young children using direct methods of assessment. In the current study, four direct measures were used: the ESB, MI, APS-RVT and the SR for age groups 2 and 3. Results showed that most children complied with the ESB and appeared to enjoy it. The only children who did not respond on most of the tasks

of the ESB showed serious delays on all other direct and indirect measures. Thus, the high level of compliance for the majority of children on this measure makes it appropriate for use with young children. In addition, since the ESB mainly assesses social engagement, noncompliance on this task is informative and might be indicative of later social communication problems (Roy & Chiat, 2014).

With regard to MI, as already discussed in section 7.3.2, most children either imitated all the items of the test or refused to participate in the task. Refusal on this task also appears to be informative and could be an indication of active interpersonal engagement difficulties. The greatest number of refusals was in the youngest age group (11.76% of age group 1). This is in line with the findings of Dohmen, Chiat and Roy (2013) who reported that refusal was observed only in the youngest age group in their typical sample and decreased with age in their specific language delay (SLD) sample.

Performance on the SR task was similar to performance on MI in terms of compliance. Most children either imitated all the items with minimal reinforcement to complete the task or did not participate at all. Nine children (8% of the sample) did not respond to the SR task. A third of those children did not comply with the MI as well. These findings run contrary to Wallan, Chiat, and Roy (2011) who did not report any refusals in their sample (A. Wallan, personal communication, October, 25, 2014). However, important differences exist between the samples in the current study and Wallan and colleagues' study. In Wallan and colleagues' study the sample mainly comprised typically developing children with no concerns about their language development, with a small group of clinical children recruited separately to compare their performance on various tasks to the typically developing children. The clinic sample in the youngest age groups (2;6-3;5 years) in Wallan and colleagues' study was very small and consisted mainly of children with speech problems. Another important factor that may have affected children's responses in this study in comparison to Wallan and colleagues' study is the elicitation context. While the present study consisted of different sociocognitive and language tasks requiring different responses from the children, the focus of Wallan and colleagues' study was verbal imitation and almost all tasks she administered required repetition of verbal stimuli. On the other hand, the current findings of non-compliance are in accord with other studies of verbal repetition. For example, Chiat and Roy (2007) reported that 6% of their

typical sample (aged 2;0-4;0) and 7% of their clinic sample (aged 2;6-3;6) did not respond to a task of word and nonword repetition.

Finally, with regard to compliance on the APS-RVT, most children in age groups 2 (2;6-2;11 years) and 3 (3;0-3;5 years) appeared to be willing to participate, although they needed encouragement to sustain attention after the first two groups of items were presented. The youngest age group (2;0-2;5 years), on the other hand, appeared to be less interested in the task and needed more encouragement to participate and complete it. Nevertheless, only 4 children from the whole sample (across all three age groups) were scored 0 because they were non-responsive or appeared to have difficulty following the instructions. While similar standardized direct assessments of vocabulary comprehension in English such as the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4; Dunn & Dunn, 2007) and the British Picture Vocabulary Test-Third Edition (BPVS3; Dunn, Dunn & Styles, 2009) are not commonly used for children as young as 2 years of age, the sensitivity of the APS-RVT in the current study in detecting differences between the three age groups and the fact that there was a range of scores in all age groups (see section 6.3.2.4) points to its validity in providing information on children's receptive vocabulary and support its usefulness in providing non-contextualized information about the vocabulary comprehension skills of children aged 2;0-3;5 years old.

To summarize, compliance rate on all the direct measures indicates that the measures are viable for children in the considered age range. Better rates of compliance were observed on the ESB and APS-RVT in comparison to the MI and SR tasks. However, in most cases non-compliance appeared to be indicative of difficulties.

7.4.5 Gender

Analyses of performance of girls and boys on the different measures used in this study revealed that mean scores were higher for girls than boys but the difference did not reach significance with the exception of the ESB on which the girls performed slightly but significantly better than the boys. Consequently, unlike the LUI in which separate norms for boys and girls were developed (O'Neill, 2009), this study found no evidence to justify separate norms. It is important to note that in an earlier study (O'Neill, 2007) with a sample smaller than the standardization study sample, O'Neill also found no difference

between the scores of boys and girls on the LUI. Thus, different findings may be found with a larger sample size in future studies.

While the gender differences found on the ESB in the current study are in contrast to results reported by Alkadhi (2009), they are in line with Chiat and Roy (2006a) who found a significant small effect of gender on the performance of girls and boys (P. Roy, personal communication, August, 27, 2014). The sample in Alkadhi (2009) consisted of only a small number of Saudi children and included only children with no concerns about their language development. Since the present study found the gender difference in the ESB occurred only in the third age group (3;0-3;5), and the difference was of a small effect size, gender was not taken into consideration in further analyses. However, the observed gender differences on the ESB are note-worthy and worth following up in future studies.

7.4.6 Representativeness of the sample

In interpreting scores in a clinical setting, it is important to consider whether recruited participants represent the general population. In this study, children were recruited from three public nurseries and one private nursery in four different areas in Riyadh, Saudi Arabia. Children attending public nurseries in Saudi usually come from varied SES backgrounds, while children attending private nurseries are mainly from middle to high SES in terms of income. The percentage of children attending nurseries in Riyadh is not known, but it is believed that more children are being enrolled in nurseries at a younger age due to increasing numbers of women joining the workforce. Nursery staffs were asked to send invitations to all children in the targeted age range who did not have hearing, visual, motor impairment, or diagnosis of ASD. However, the researcher did not have access to the number of children in each nursery, number of sent invitations and whether staff did approach all parents or were biased in sending invitations to parents who were more likely to respond. The researcher also did not have access to the number of parents who declined and their demographics. Children whose parents signed the consent forms were included. Based on the educational level of the parents, it was determined that most of the sample were from middle class backgrounds according to parental education. As mentioned in section 5.1.5, it was difficult to determine how the sample compares to the Saudi population due to lack of statistics on these factors. Nevertheless, as argued earlier, the sample is probably representative of the children encountered in a clinical setting:

based on personal experience, middle class educated parents of children with language or communication delay are the ones who are most likely to seek help from speech and language therapy clinics. An important point to bear in mind is that although the SES for the sample was determined based on parents' education, the educational level may not reflect the parents' economic level as the educational system in Saudi Arabia offers free education from school level to university level. In fact government universities which offer free education up to doctorate level were established in Saudi before the establishment of private universities.

7.5 Clinical Implications

A main motivation for undertaking this project was the lack of assessment tools and normative data in Saudi Arabia. This study has produced a substantial set of data on a wide-ranging battery of assessments that can be used for clinical comparison. Although the data collected on the newly developed or adapted measures was from children in Riyadh speaking the Najdi dialect, most of the measures can be used in other regions in Saudi with little or no change. For example, the sociocognitive measures (ESB and MI), being essentially non-verbal, can be used with children from different dialectal backgrounds. Furthermore, the APS-RVT can be easily adapted to other dialects of Saudi Arabic as there are very few dialectal differences between regions in Saudi on the items used in the APS-RVT. The ARA-LUI test instructions were adapted from English to standard Arabic. Since the communicative functions that are assessed in the LUI did not include items addressing quality and manner, which may be more influenced by culture (O'Neill, 2014); the adapted version appears to be applicable to Arabic speaking children from different dialectal and cultural backgrounds. However, future research may further investigate the appropriateness of items that assessed aspects of semantics and syntax such as the parts that assessed use of different word classes, mental state terms, modals, time indicators and conjunctions. In adapting these parts it was not possible to choose items based on frequency or age of acquisition due to lack of research on Arabic language acquisition or frequency data.

Of great clinical significance is that the current findings suggest that the combination of measures used in this study have the potential to guide clinicians across different cultures and linguistic backgrounds in evaluating risk status and prioritizing caseloads. Identifying

language impairment is not an easy matter even for English-speaking children for whom numerous standardized measures of language have long been available. Results of the current study suggested that severity of impairment was associated with pervasiveness, with children showing impairment across the language and sociocognitive measures attaining lower scores than children performing low only on language measures. Thus, a child who shows delays on language and sociocognitive measures may require immediate intervention services. Conversely, a child with delays only on language measures, with normal sociocognitive skills, might be at a lower risk. Eligibility for intervention in this group should consider the extent of the delay according to direct and indirect measures and the degree of impairment on these. For example, children performing low on all language measures may be recommended for intervention services, whereas children with severe delays only on one language measure or those with minor delays across more than one language measure may be classified as the "watch and see" group as designated by Paul and Roth (2011). Children falling in this category may be followed every 3-6 months to monitor their progress and re-evaluate their eligibility for services.

Moreover, the battery used in this project included measures that enable the identification of a language delay (the ARA-LUI and the APS-RVT), as well as measures that aid in the identification of deficits underlying the delay (the ESB and MI). The battery can therefore contribute not only to the identification of children with language and/or communication delays, but also to planning intervention, by identifying deficits in underlying skills as targets for intervention. In addition, it provides information on the child's abilities from different sources using parental reports (ARA-LUI) and direct methods of assessment. This is in line with recent views on assessing children with language delays that advocate "plac[ing] less reliance on simplistic models of discrepancy and mak[ing] greater attempts to characterize the child's performance on different tasks and situations resulting in a profile of skills and needs" (Dockrell & Marshall, 2015, p. 4).

Finally, it must be remembered that although the results obtained on the various measures can provide a reference allowing objective identification of children performing at the lower end of the range in Saudi, they cannot be treated as norms as this requires a larger number of participants who are truly representative of the population (Rust & Golombok, 2009) and interpretation of present scores should consider issues of sampling addressed in the previous sections. Furthermore, to confirm informativeness of the battery of measures,

it is essential for future research to conduct longitudinal studies that assess longer term impact and individual functioning outcome based on different levels of performance on the different measures.

7.6 Conclusion and Future Directions

A key aim of this study was to develop a battery of measures to assess young Saudi children using measures reported to be informative about concurrent strengths and difficulties and later outcomes. The unique range of measures that were either adapted or newly developed were found to be reliable, valid and all but one were age sensitive, therefore show potential in identifying children with language delays. A substantial set of data has been produced on the different measures which can be used as a reference to which clinically referred Saudi children can be compared, enabling objective identification of low performance. This is of great clinical value given that there are almost no available tools in Arabic for identifying these children and forms the first step towards developing standardized measures. Moreover, by including measures that assess skills underlying language the battery can guide intervention by identifying the nature of children's impairment.

In addition, this is the first study to examine relations between sociocognitive skills and language in an Arabic language culture, with most previous studies conducted in English-speaking populations. Results were remarkably similar to those in Western cultures, showing that sociocognitive skills are important predictors to language outcome that may transcend specific language and culture. The present findings contribute to the evidence of trajectories of language acquisition and point to the importance of including measures of sociocognition when assessing young children with language delay across different cultures. Given that the sociocognitive measures used in the current study are essentially nonverbal, they have the potential of being informative when used with children from different cultures and speaking different languages. They may be particularly valuable where there are no assessments in the child's language and/or the assessor does not speak the child's language. Due to practical concerns, only concurrent predictiveness was investigated and following up children was beyond the scope of this project. Future research should evaluate the measures with a larger, more representative sample with an in-depth consideration of what representativeness means in the Saudi context; examine

relations between performance on the different measures and later outcome; administer the battery of measures to clinically referred children and assess the measures' diagnostic accuracy and predictive value; and examine relations between measures in clinically referred children to determine implications of different profiles. This will greatly extend evidence on the informativeness of the measures and our understanding of the developmental trajectories of language.

APPENDICES

APPENDIX A: Sociocognitive Questionnaire Items and Rationale

	Item	Tests	Rationale
1.	Responds to his/her name when called	M-CHAT[a], CSBS-DP[b], GARS[c], FYI[d]	Failure to respond to name is one of the early indicators of autism (Robins, Fein, Barton & Green, 2001; Wetherby et al., 2004)
2.	Looks at a toy across the room when you are looking at it	M-CHAT, CSBS-DP	Gaze following predicts language abilities (Brooks & Meltzoff, 2008)
3.	Looks at a toy across the room when you point to it	M-CHAT, CSBS-DP, ESAT[e], FYI	Point following predicts language abilities (Mundy et al., 2007)
4.	Looks at you when he/she is playing with a toy to see if you are watching him	CSBS-DP	Significant correlations were reported between IJA and language abilities (Mundy et al., 2007)
5.	Shows you a toy or an object just to get you to notice it, not to ask you to do anything with it	CSBS-DP, SSI[f], M-CHAT	Lack of showing distinguished children with ASD from TD children and children with DD (Wetherby et al., 2004)
6.	Smiles when you smile	M-CHAT	Children with ASD are less attentive to others' emotions (Beeger, Koot, Rieffe, Terwogt, & Stegge, 2007)
7.	Looks at your face when faced with something unfamiliar to check your reaction	M-CHAT, FYI	Social referencing deficits are characteristic of children with autism (Bacon, Fein, Morris, Waterhouse, & Allen, 1998)
8.	Tries to provide comfort when someone is hurt or sad	SCQ[g]	Infants later diagnosed with ASD are impaired in attention to distress (Hutman et al., 2010) Children with autism show low levels of responsiveness in comparison to DD and TC (Dawson et al., 2004)
9.	Engages in pretend play e.g. pretends to talk on the phone, feed a stuffed animal, or putt a doll to sleep	M-CHAT, CSBS-DP, ESAT	Pretend play competence are strongly associated with later language comprehension (Laakso, Poikkeus, Eklund, & Lyytinen, 1999)
10.	Uses an object in a pretend way, as if it is a different object e.g. using a banana as a phone, using a shell as a cup	ROSSETTI[h], CSBS-DP	Symbolic play correlated with receptive and expressive language (Lewis, Boucher, Lupton, & Watson, 2000).
11.	Imitates other people's actions while playing	GARS, FYI	Motor imitation is related to concurrent language abilities (Sigman & Ungerer, 1984) and predictive of later development (McDuffie, Yoder, Stone, 2005)
12.	Does things just to get you to laugh	CSBS-DP	Children with autism show limited production of humor (Reddy, Williams, & Vaughn, 2002; St. James & Tager-Flusberg, 1994)

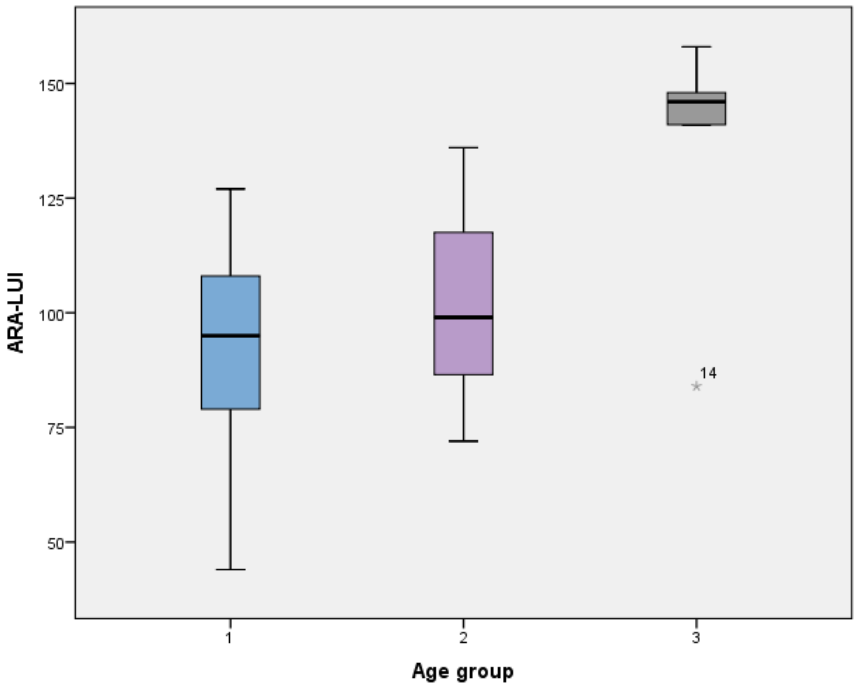
Table continued overleaf

13. Overactive, cannot stay still for long	SDQ[i], ASQ-SE[j]	High correlations between pragmatic competence and hyperactivity (Ketelaars, Cuperus, Jansonius & Verhoeven, 2010)
14. Prefers to play alone	SDQ	High levels of withdrawal were reported in children with SLI (Hart, Fujiki, Brinton, & Hart, 2004; Stanton-Chapman, Justice, Skibbe & Grant, 2007)
15. Shows interest in other children	M-CHAT, ESAT, SCQ	Late talkers scored lower on nonverbal socialization items on VABS[k] in comparison to TC e.g. interest in children (Paul, Looney, & Dahm, 1991)
16. Points using his index finger to ask for an object	M-CHAT, CSBS-DP, AEPS[l]	Pointing predicts language scores (Mundy et al., 2007; Brooks & Meltzoff, 2008)
17. Interacts with adults better than children	SDQ	Preschoolers with SLI and speech impairment are more likely to interact with adults than with peers (Rice, Shell, & Hadley, 1991)
18. Waves to greet	CSBS-DP	Conventional gestures like waving are predictive of later expressive vocabulary (McCathren, Yoder, & Warren, 2000)

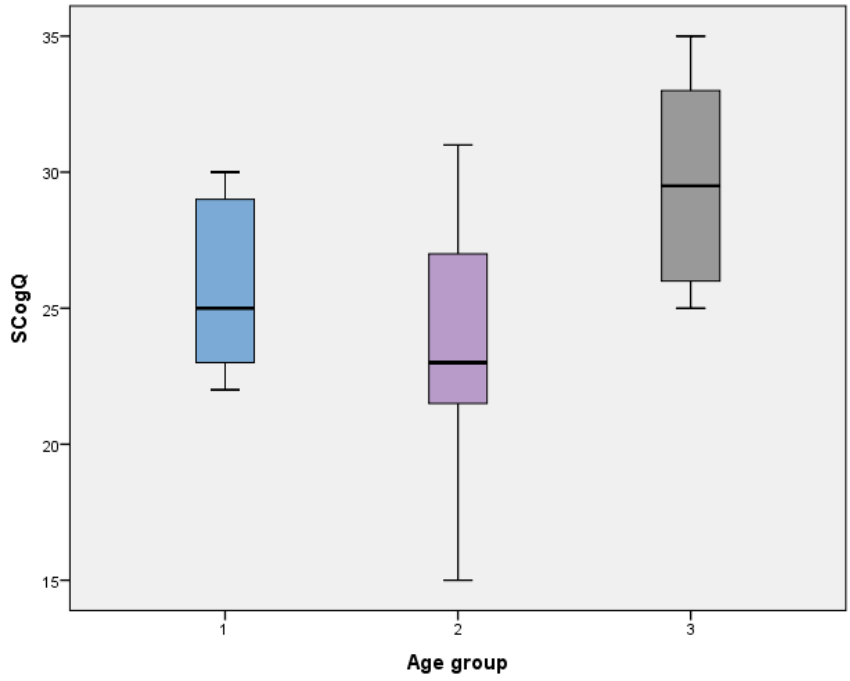
[a] M-CHAT: Modified checklist for Autism in Toddlers (Robins, Fein, Barton, & Green, 2001).
[b] CSBS-DP: Communication and Symbolic Behavior Scales Developmental Profile (Wetherby & Prizant, 2002).
[c] GARS: Gilliam Autism Rating Scale (Gilliam, 1995).
[d] FYI: First Year Inventory (Baranek, Watson, Crais, & Reznick, 2003).
[e] ESAT: Early Screening of Autistic Traits Questionnaire (Swinkels et al., 2006).
[f] SSI: Ghuman-Folstein Screen for Social Interaction (Ghuman, Freund, Reiss, Serwint, & Folstein, 1998).
[g] SCQ: Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003).
[h] ROSSETTI: The Rossetti Infant-Toddler language Scale (Rossetti, 1990).
[i] SDQ: Strengths and Difficulties Questionnaire (Goodman, 1997).
[j] ASQ-SE: Ages and Stages Questionnaire: Social-Emotional (Squires, Bricker, Twombly, 2002).
[k] VABS: Vineland Adaptive Behavior Scales (Sparrow et al., 1984).
[l] AEPS: Assessment, Evaluation, and Programming System for infants and Children (Cripe & Slentz, 1993).

APPENDIX B: First Pilot Results

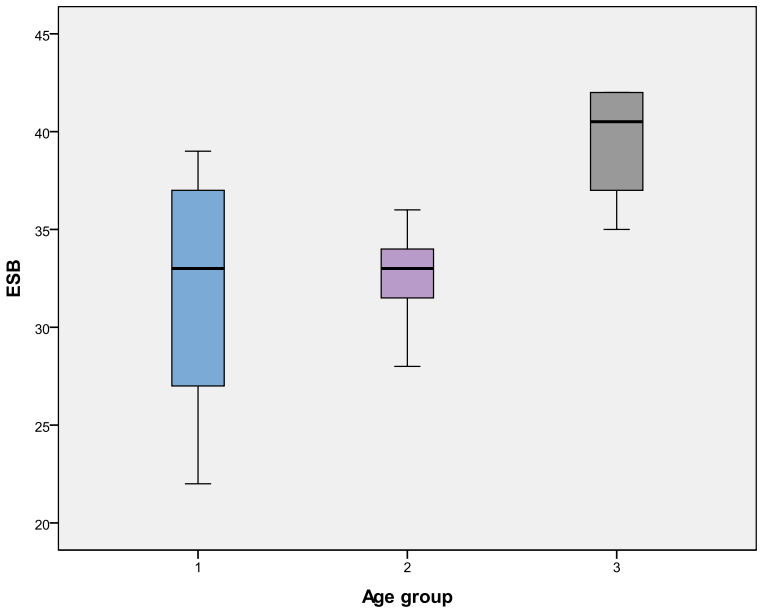
Box plots of the scores of the three age groups (n = 19) on the ARA-LUI



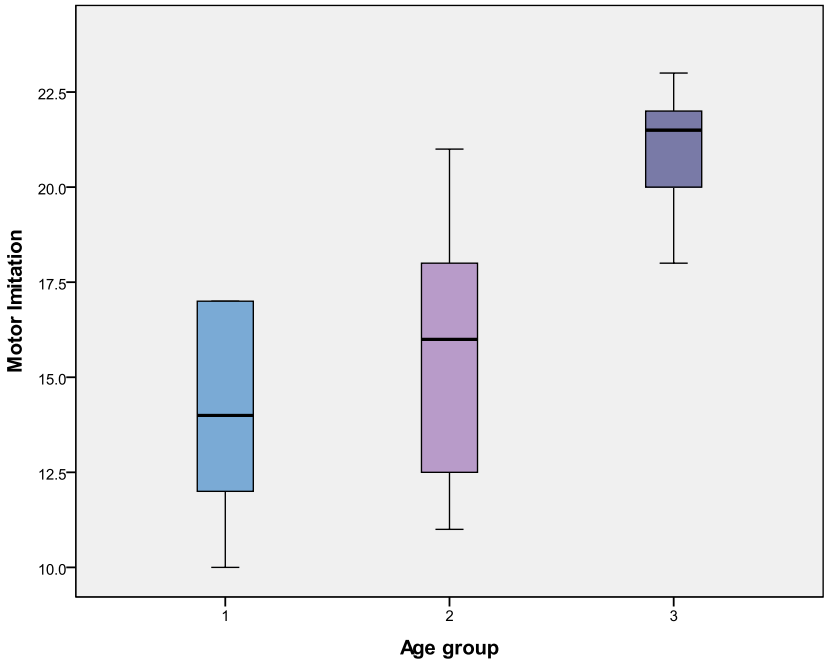
Box plots of the scores of the three age groups (n = 19) on the SCogQ



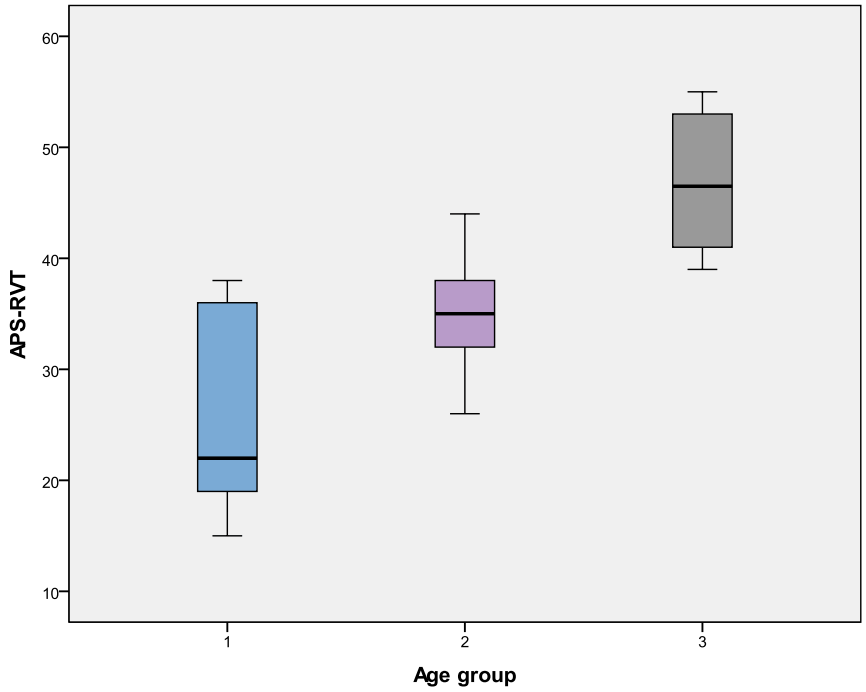
Box plots of the scores of the three age groups (n = 19) on the ESB



Box plots of the scores of the three age groups (n = 19) on the Motor Imitation task



Box plots of the scores of the three age groups (n = 19) on the APS-RVT



Partial correlations between measures controlling for age in months (n =19)

	ARA-LUI	SCogQ	ESB	MI
SCogQ	.39 [-.04, .80]			
ESB	.29 [-.21, .70]	.01 [-.42, .53]		
MI	.62** [.32, .84]	.48* [-.23, .86]	.23 [-.22, .60]	
ASP-RVT	.59* [.04, .89]	-.08 [.-58, .50]	.54* [.08, .86]	.59* [.04, .89]

*p < .05, **p < .01, BCa bootstrap 95% CIs reported in [ ]

APPENDIX C: Items of the Arabic Preschool-Receptive Vocabulary Test

Group 1		Group 2		Group 3		Group 4	
1.	ʕeɪn eye	13.	meknesa broom	25.	ħɑːmɪdˤ sour	37.	bajdˤɑːwɪː oval
2.	jɪrkɪdˤ run	14.	nemer tiger	26.	jeʃəg tear	38.	mafdʒuːʕa surprised
3.	seɪkal bike	15.	sˤabuːn soap	27.	jeħɑːsɪb pay	39.	ħemɑːr waħʃɪː zebra
4.	bessa cat	16	tɪqra read	28.	taʕabɑːna exhausted	40.	χaʃabɪː wooden
5.	bɑːsˤ bus	17.	dʊktɔːr doctor	29.	ʃamʕa candle	41.	ħalazɔːn snail
6.	beɪjbɪː baby	18.	jenɪtˤ jump	30.	jebuːs kiss	42.	teʃuːf see
7.	festɑːn dress	19.	ʔsˤbaʕ finger	31.	sˤeffɑːra whistle	43.	teχredʒ leave
8.	ʕadʒɪːna dough	20.	jeħfer dig	32.	teχɑːnɪg scold	44.	jeχarreb destroy
9.	bagara cow	21.	warda flower	33.	θʊʕbɑːn snake	45.	tˤabla drum
10.	beɪt house	22.	jeblaʕ gulp	34.	makɪːnat χejɑːtˤa sewing Machine	46.	mesmɑːr nail
11.	jɪʃrab drink	23.	daʔɪra circle	35.	tˤɑːwuːs peacock	47.	tegtˤef pick
12.	bʊrtʊqɑːl orange	24.	manfuːχ blown	36.	mʊʔddab polite	48.	ħemɑːr donkey

Table continued overleaf

Group 5		Group 6		Group 7	
49.	jegɪːs measure	61.	ʤɔːz ʔɪlhend coconut	73.	qalʕah castle
50.	mekaʃʃer frawning	62.	jetsllag climb	74.	mʊmarreðˁah nurse
51.	getɑːr guitar	63.	teleskɔːb telescope	75.	taʕɑːwn cooperation
52.	kanʁar kangaroo	64.	sˁajjɑːd hunter	76.	ħaʃiːʃ grass
53.	jegɑːbɪl meet	65.	faχuːr proud	77.	jefattɪʃ search
54.	χarbɑːn destroyed	66	tefħasˁ examine	78.	ʔħfɑːd grandchildren
55.	tasˁɑːdʊm collision	67.	qamħ wheat	79.	mʊhandɪs engineer
56.	rʊkbah knee	68.	bʊrʤ tower	80.	bɑːzella peas
57.	ʕeʃ nest	69.	faras ɪlbħr seahorse	81.	χajɑːlɪː fictional
58.	meʃawwek thorny	70.	kaʕab heel	82.	gʊfl lock
59.	jerfes kick	71.	mʊftarɪs predator	83.	hʊdhʊd hoopoe
60.	deɪnasˁɔːr dinosaur	72.	ðeɪl tail	84.	ʃewajjah few

APPENDIX D: Head of Nursery Invitation, Parent's Information Sheet, Consent Forms and Demographic Questionnaire

CITY UNIVERSITY LONDON

Re: An investigation of relations between sociocognitive skills, motor imitation and language skills in young Saudi children.

Dear Head of Nursery

I am a qualified speech language pathologists with experience in assessing and working with young children. Currently, I am undertaking my PhD in City University London. As part of my research project I am investigating the performance of young Saudi children with no known developmental disorders on sociocognitive tasks and motor imitation tasks which are known to be informative about children's language and communication abilities.

The purpose of this study is to find out how Saudi children perform on these tests so that professionals in Saudi Arabia can use them for clinical assessment. This will help with early detection of children with communication disorders, and early intervention to help these children.

For my study, I aim to see 180 children aged 2-3½ years, and I am writing to ask if you would be willing to take part. If so, I would ask you to distribute information sheets and consent forms to parents of children who meet the criteria for participating in the study:

- First language is Arabic.
- No hearing loss, or significant visual impairment.
- No motor impairment or congenital problems.
- Absence of developmental disorder, autism, or a neurological disorder of known etiology.

Only children whose parents give signed consent will be included in the study. We would need to carry out the assessment tasks in a quiet room at your nursery. Each child will be seen individually for two session each lasting up to 30 minutes, if possible, with the child's parent or teacher attending the session. I will discuss timing of the sessions with their teacher to ensure any disruption is minimal.

After establishing rapport with the child, the researcher will introduce the assessment tasks which take the form of games. During these games the researcher will observe the child's reaction to expressions of emotion, measure their capacity to share attention by monitoring their gaze, assess their understanding of gestures and objects used as symbols, and test the child's ability to imitate different actions modeled by the experimenter. These non-verbal skills are important for language development. Then two language assessments will be administered: a sentence repetition task where the child will be asked to repeat sentences after the researcher and a receptive vocabulary task where the child will be asked to point to pictures named by the researcher.

I will only include children who are willing to join the researcher and participate in the activities. During the assessments, if the child was unwilling to continue, I will stop the session and take the child back to an appropriate member of staff.

CITY UNIVERSITY LONDON

In addition to assessing the children, I will ask parents to complete a questionnaire to provide demographic information, a report on their children's use of language, and a questionnaire on their observations of their child's communicative and social skills.
All information provided will be confidential. Each child participating in this study will be given a number and the results will be anonymously analyzed as they will include only the children's numbers. The list of names and numbers will be kept on a password protected electronic document. The list of children's names and identity numbers will be erased 5 years after the end of the project. Only the anonymous records of data will be kept.

Thank you for giving this your consideration.

If you have any further questions or would like to discuss any part of the study, please do not hesitate to contact us.

Investigator
Aseel Alkadhi
[redacted]

Project supervisors
Professor Shula Chiat
[redacted]

Professor Penny Roy
[redacted]

City University London
Northampton Square
London
EC1V 0HB

If there is an aspect of the study which concerns you, you may make a complaint. City University London has established a complaints procedure via the Secretary to the Research Ethics Committee. To complain about the study, you need to phone [redacted] You can then ask to speak to the Secretary of the Ethics Committee and inform them that the name of the project is: An investigation of relations between sociocognitive skills, motor imitation and language skills in young Saudi children.
You could also write to the Secretary at:
Anna Ramberg, Secretary to Senate Ethics Committee, CRIDO, City University, Northampton Square, London, EC1V 0HB
Email: [redacted]

CITY UNIVERSITY LONDON

Informed Consent Form for Head of Nursery

Project title: An investigation of relations between sociocognitive skills and language skills in young Saudi children

Supervisors: Professor Shula Chiat
Professor Penny Roy
Department of Language and Communication Science
City University London, Northampton Square, London EC1V 0HB

Investigator: Aseel AlKadhi E-mail: [redacted]

1. I confirm that I have read and I understand the explanatory statement for the above City University project and have been offered the opportunity to ask questions.

2. I agree for my nursery to take part in the above study which means that:

- I allow Aseel AlKadhi to assess children in my nursery whose parents agree to their participation on the sociocognitive tasks, motor imitation tasks and language tests. I also allow some of the assessments to be audio recorded.
- I understand that all information collected in the research is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party. No identifiable personal data will be published. The identifiable data will not be shared with any other organization. Each child will be given a number and the list of names and numbers will be kept on a password protected electronic document. Score sheets and databases will be anonymous as they will include only the child's number. The list of children's names and identity numbers will be erased 5 years after the end of the project. Only the anonymous records of data will be kept. The school's identity will be protected and will not appear in the report.
- I understand that taking part is voluntary and that my nursery is free to withdraw from the study at any time, without giving any reason.

Name of Nursery:…….. .... Date:……

Name of Nursery head:……..... Signature:……

If there is an aspect of the study which concerns you, you may make a complaint. City University has established a complaints procedure via the Secretary to the Research Ethics Committee. To complain about the study, you need to phone [redacted] You can then ask to speak to the Secretary of the Ethics Committee and inform them that the name of the project is:
An investigation of relations between sociocognitive skills, motor imitation and language skills in young Saudi children.
You can also write to the Secretary at:

Anna Ramberg
Secretary to Senate Research Ethics Committee
CRIDO
City University
Northampton Square
London
EC1V 0HB
Email: [redacted]

Or you can report your complaint to the head of the nursery.

Further Information

If you have any further concerns or questions, please do not hesitate to contact us.

Project Supervisors:
Professor Shula Chiat [redacted]

Professor Penny Roy [redacted]

City University London
Northampton Square
London
EC1V 0HB

Investigator:
Aseel AlKadhi [redacted]

Secretary of Ethics Committee:
Anna Ramberg [redacted]

CITY UNIVERSITY LONDON

An invitation for your child to take part in a research project

An investigation of relations between sociocognitive skills, motor imitation and language skills in young Saudi children

ABOUT THE RESEARCH PROJECT

We are researchers from City University London interested in finding ways of identifying children at risk of language and communication problems.

A number of studies have reported that children's sociocognitive skills and motor imitation skills are informative about children's later language and communication abilities. We are investigating the performance of young Saudi children with no known developmental disorders on an assessment that has been developed in the UK to assess pre-school children's social communication. In addition children's motor imitation will be studied and relations between children's performance on the sociocognitive tasks, motor imitation tasks and language skills will be explored.
We need to know how typically developing Saudi children perform on these tests before professionals can use them to identify children at risk of communication problems.

WHAT IS INVOLVED

We aim to see 180 children whose parents agree to their participation, and who are themselves willing to participate. A researcher will see each child individually in the nursery for two sessions each lasting up to 30 minutes. After establishing rapport with the child, the researcher will introduce each of the tasks which take the form of games. During these games, she will observe the child's reaction to expressions of emotion, measure their capacity to share attention by monitoring their gaze, assess their understanding of gestures and objects used as symbols, and assess their imitation of actions modeled by the researcher. She will also carry out two language assessments. In one, she will ask the child to repeat sentences after the researcher, and in the other, to point to pictures that she names.

We will also send you three questionnaires to complete. These will ask you about your child's background, and your observations of your child's use of language, and social and communication skills.

If you are willing for your child to participate in the project, we would be grateful if you would fill in the attached consent form and return it to the head of the nursery within a week.

We will only include children if they agree to join the researcher and participate in the activities. Children have been found to enjoy these activities, but if your child does not want to go on, we will stop the session and take the child back to an appropriate member of nursery staff.

Each child participating in the assessments will be given a number, and the list of names and numbers will be kept on a password protected electronic document. The assessment results will be anonymous as they will include only the children's numbers. Five years after the end of the project, the list of children's names and identity numbers will be erased. Only the anonymous records of data will be kept.

If the assessments raise any concerns regarding your child's development, we will discuss these concerns with you and with the child's teacher if you agree to this.

Your child does not have to take part in this study if you do not want them to, and even if you agree to your child taking part, you may withdraw them at any time without having to give a reason.

Many thanks for giving this your consideration.

CITY UNIVERSITY LONDON

Informed Consent Form for Parents/Guardians of Project Participants

Project title: An investigation of relations between sociocognitive skills, motor imitation, and language skills in young Saudi children

Supervisors: Professor Shula Chiat
Professor Penny Roy
Department of Language and Communication Science
City University London, Northampton Square, London EC1V 0HB

Investigator: Aseel AlKadhi E-mail: [redacted]

I agree that my child…… (full name of child) for whom I am a guardian may take part in the above City University research project. The project has been explained to me, and I have read the Explanatory Statement, which I may keep for my records.

I understand that agreeing to take part means that I am willing to complete the questionnaires and to allow Aseel AlKadhi to assess my child on the sociocognitive tasks, motor imitation tasks and language tests. I also allow some of the assessments to be audio recorded.
This information will be held and processed for the data analysis of the project.

I understand that any information ……(full name of child) provided is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party. No identifiable personal data will be published. The identifiable data will not be shared with any other organization. Each child will be given a number and the list of names and numbers will be kept on a password protected electronic document. Score sheets and databases will be anonymous as they will include only the child's number. The list of children's names and identity numbers will be erased 5 years after the end of the project. Only the anonymous records of data will be kept.

I also understand that my child's participation in the project is voluntary, and that I can withdraw my child from the study without being penalized or disadvantaged in any way.

CITY UNIVERSITY LONDON

In case the assessments raise some concerns regarding my child's development I allow the researcher to discuss these issues with my child's teacher yes☐ no☐

Participant's Name:…… (please print)

Participant's Age:……

Parent's/ Guardian's Name:……

Your relationship to participant:……

Signature of Parent/Guardian:……..

Date:……

CITY UNIVERSITY LONDON

An investigation of sociocognitive skills in young Saudi children
Parental questionnaire

[1] Child's name..... Boy Girl *(please circle)*

[2] Today's date..... Your child's date of birth.....

[3] Number of Siblings..... Child's order in the family.....

[4] What is your child's language background? *(please circle one)*
Home:
a. Arabic language only
b. Arabic and other languages Please indicate languages used.....
Nursery/ School:
a. Arabic language only
b. Arabic and other languages Please indicate languages used.....

[5] Do you have any concerns about your child's hearing? *(please circle one)*
Yes No

If yes, please specify:

[6] Does your child have a medical or neurological diagnosis? *(please circle one)*

Yes No

If yes, please specify:

[7] The mother's education level
a. High school degree b.Diploma c. Bachelor degree d. Postgraduate degree
e. Other (please specify).....

Occupation of mother.....

[8] The father's education level
a. High school degree b.Diploma c. Bachelor degree d. Postgraduate degree
e. Other (please specify).....

Occupation of father.....

APPENDIX E: Percentage of Saudi Employees by Educational Level

Percentage Distribution of Saudis employees by (educational level, age group, Occupation and economic activity)	2009	2008	2007	2006
Illiterates	**3.9**	**3.9**	**4.8**	**5.3**
Male	4.1	4.1	5.0	5.6
Female	2.1	2.3	3.6	3.9
Below Secondary	**31.6**	**33.7**	**33.8**	**35.7**
Male	35.5	38.1	38.5	41.0
Female	5.6	5.5	4.7	4.5
Secondary	**29.0**	**37.4**	**24.3**	**23.1**
Male	31.8	28.8	26.4	24.8
Female	10.4	92.8	11.6	12.8
Diploma	**9.1**	**8.7**	**10.0**	**9.0**
Male	7.9	7.5	8.1	7.3
Female	17.2	16.9	22.0	18.7
Academic and Over	**26.5**	**26.9**	**27.1**	**26.9**
Male	20.7	21.6	22.0	21.3
Female	64.6	61.1	58.1	60.1

Information derived from statistical data and indicators on Saudi human resources for 2006-2009, retrieved from http://www.cdsi.gov.sa/

APPENDIX F: Sentence Repetition Test

1									
ʃa:f	tense	mHammad	ʔXwan	pat	-a	f-	il-	madras	-a
saw	pf	prop.n	brother	plural	his	in	the-	school	-fsg

2									
ʔil-	walad	saʔal	tense	ṣidi:g	-a	ʕan	ʔil-	Hafl	-a
the-	boy	ask	pf	friend	-his	about	the-	party	-fsg

3										
Haṭ	-at	tense	ʔil-	bint	daftar	-ha:	ʕala:	iṭ-	ṭa:wl	-a
put	-pf3fsg	pf	the-	girl	notebook	-her	on	the-	table	-fsg

4										
ʤu:d	ʃar	tense	-at	haða:	pattern	il-	galam	min	ʔil-	maHal
prop-n	buy	pf	-pf3fsg	this	msg	the-	pen	from	the-	store

5										
ʤalas	tense	ʔil-	walad	ʔiṭ-	ṭwi:l	ʕala	haða:	patter	il-	kursi:
sit	pf	the-	boy	the	tall	on	this	msg	the-	chair

6											
maha:	ka:n	tense	-at	ti-	sbaH	tense	maʕ	Xal	-ha:	il-	kibi:r
PropN	was	pf	-pf3fsg	imp3fsg-	swim	imp	with	brother	-her	the	big

Sentence Repetition Test (A. Wallan, personal communication, October, 9, 2011)

7											
nu:ra:	ʁasal	-at	tense	-ha:	bi-	ʔil-	mo:j	-a	w-	is-	ʂabu:n
prop-n	wash	-pf3fsg	pf	-her	with	the-	water	-fsg	and-	the-	soap

8											
ʤar	-at	tense	ʔil-	biss	-a	is-	so:da	pattern	wara:	il-	fi:l
run	-pf3fsg	pf	the-	cat	-fsg	the-	black	fsg	after	the-	elephant

9											
ʔil-	walad	iʂ-	ʂi:ʁi:r	ʔaXað	tense	haði:	pattern	il-	ku:r	-a	minn	-a
the-	boy	the-	small	take	pf	this	fsg	the-	ball	-fsg	from	-him

10												
ji-	gra:	tense	hu:	pattern	kita:b	ʕan	ʔid-	di:k	w-	in-	naml	-a
imp3msg-	read	imp	pro	msg	book	about	the-	hen	and	the-	ant	-fsg

11												
ka:n	tense	na:jif	ji-	rkiðˌ	tense	f-	il-	Hadig	-a	maʕ	ʔaʂHa:b	pat	-a
was	pf	prop.n	imp 3sg-	run	imp	in-	the-	garden	-f	with	friend	plural	-his

Sentence Repetition Test (A. Wallan, personal communication, October, 9, 2011)

12													
hu:	pattern	ji-	Hib	tense	ji-	ʃrab	tense	ʔil-	Hali:b	bi-	il-	fara:wl	-a
he	msg	imp3 msg-	love	imp	imp3 msg-	drink	imp	the-	milk	with-	the-	straw-berry	fsg

13													
ti-	ʕi:ʃ	tense	ʔiz-	zara:f	-a	il-	Hilw	-a	fi:	haði:	pattern	il-	ʁa:b	-a
imp 3fsg-	live	imp	the-	girrafe	-fsg	the	pretty	-fsg	in	this	fsg	the-	forest	-fsg

14														
hij	pattern	ti-	Hib	tense	ti-	ʁab	tense	b-	ʕaru:s	-at	-ha:	il-	ʤidi:d	-a
he	fsg	imp 3fsg-	love	imp	imp 3fsg-	play	imp	with-	doll	-fsg	-her	the-	new	fsg

Sentence Repetition Test (A. Wallan, personal communication, October, 9, 2011)

	lexical		grammatical		CELF	
1	4		6		3	
2	4		6		3	
3	4		7		3	
4	4		7		3	
5	4		7		3	
6	4		8		3	
7	4		8		3	
8	4		8		3	
9	4		9		3	
10	4		9		3	
11	4		10		3	
12	4		10		3	
13	4		11		3	
14	4		11		3	
	56		117		42	

Sentence Repetition Test (A. Wallan, personal communication, October, 9, 2011)

APPENDIX G: SCogQ and Direct Assessments Record Forms

The Sociocognitive Questionnaire

Child's name:	Child's Date of Birth:	Today's Date:
Name of person completing the questionnaire:		Relationship to child:

Thank you for taking the time to complete this questionnaire. Please answer **All** the questions by putting √ in the box that best describes your observations of your child's typical behavior.

	Often	Sometimes	Never
1. Responds to his/her name when called.	☐	☐	☐
2. Looks at a toy across the room when you are looking at it.	☐	☐	☐
3. Looks at a toy across the room when you point to it.	☐	☐	☐
4. Looks at you when he/she is playing with a toy to see if you are watching him.	☐	☐	☐
5. Shows you a toy or an object just to get you notice it, not to ask you to do anything with it.	☐	☐	☐
6. Smile when you smile.	☐	☐	☐
7. Looks at your face when faced with something unfamiliar to check your reaction.	☐	☐	☐
8. Tries to provide comfort when someone is hurt or sad.	☐	☐	☐
9. Engages in pretend play e.g. pretends to talk on the phone, feed a stuffed animal, or putt a doll to sleep.	☐	☐	☐
10. Uses an object in a pretend way, as if it is a different object e.g. using a banana as a phone, using a shell as a cup.	☐	☐	☐
11. Imitates other people's actions while playing.	☐	☐	☐
12. Do things just to get you to laugh.	☐	☐	☐
13. Overactive, cannot stay still for long.	☐	☐	☐
14. Prefers to play alone.	☐	☐	☐
15. Shows interest in other children.	☐	☐	☐
16. Points using his index finger to ask for an object.	☐	☐	☐
17. Interacts with adults better than children.	☐	☐	☐
18. Waves to greet.	☐	☐	☐

ESB scoring sheet

No:____

Child Name: _____ Test Date:_____ Birth Date:_____

1. Social Responsiveness Assessment:

Scoring:

Responsiveness to the expression of emotion is scored in terms of looks to the assessor's face as the emotion is expressed. Points are awarded as follows:

2-The child looks at the assessor's face for at least 2 seconds.
1-The child looks briefly or fleetingly at the assessor's face.
0-The child does not look at the assessor's face at all.

Maximum total score for social responsiveness = 12

SCORE CHART FOR SOCIAL RESPONSIVENESS ASSESSMENT				
	Looks ≥ 2 seconds Score = 2	Fleeting look Score = 1	No look Score = 0	
hurt				
surprise				
anger				
fear				
distraction				
achievement				
Total				

Chiat and Roy (2006b)

2. Joint Attention Assessment:

Scoring:

Measuring gaze switch:

Either

Look from egg to adult is shaking egg (i.e. before opening egg)

Or

Look from toy to adult after the egg has been opened, while showing contents = 1

No look in either of the above conditions = 0

Measuring gaze monitoring:

Look following adult's gaze switch and verbal statement = 2

Look following adult's point and repeated verbal statement = 1

No look in either of the above conditions = 0

Maximum total score for joint attention = 18

Score chart

SCORE CHART FOR JOINT ATTENTION ASSESSMENT				
	Gaze switch	**Gaze monitoring**		
	While shaking egg Or While showing contents of egg Score = 1	After adult's gaze+statement Score = 2	After adult's point+statement Score = 1	
person				
hat				
candle				
bag				
tiger				
ring				
Total				

Chiat and Roy (2006b)

3. Symbolic Comprehension Assessment

Scoring:

One mark is awarded for each object that is correctly selected in each condition.
Maximum score = 18

Score Chart

SCORE CHART FOR SYMBOLIC COMPREHENSION ASSESSMENT							
Practice							
bag		car		flower		fork	
Assessment							
Gesture		Miniature		Substitute objects			
toothbrush		teddy		hat			
comb		brush		telephone			
hammer		book		crayon			
bottle		shoe		plate			
sock		spoon		soap			
scissors		t-shirt		ball			
Total							

Comments: _____
_____
_____

Chiat and Roy (2006b)

Motor Imitation Scoring Sheet No:_____

Child name: _____ Test Date:_____ Birth Date:_____

Practice					
Touch earlobe		Grab nose			
Score Chart for **Postures** Imitation					
	Accurate response = 2	Partial success = 1	Failure = 0	No response = 0	
Touch back of head					
Interlink fingers					
Pat elbow with one hand					
Bend index finger					
Wiggle a thumb					
Open one fist					
Total					
Score Chart for **Gestures** Imitation					
Pour					
Fly a plane					
Stir					
Turn the steering wheel					
Throw a ball					
Pull a rope					
Total					

Motor Imitation total = ____

Items of the Arabic Preschool-Receptive Vocabulary Test

No._____

Name _____ Date_____ Birth Date _____

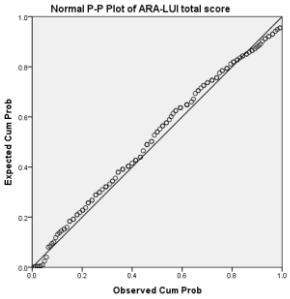
Group 1				Group 2				Group 3				Group 4			
1.	ʕeɪn eye	3		13.	meknesa broom	1		25.	ħɑ:mɪdˁ sour	2		37.	bajdˁɑ:wɪ: oval	2	
2.	jɪrkɪdˁ run	2		14.	nemer tiger	2		26.	jeʃəg tear	3		38.	mafʤu:ʕa surprised	1	
3.	seɪkal bike	3		15.	sˁabu:n soap	4		27.	jeħɑ:sɪb pay	1		39.	ħemɑ:r waħʃi: zebra	4	
4.	bessa cat	1		16.	tɪqra read	1		28.	taʕabɑ:na exhausted	4		40.	χaʃabɪ: wooden	2	
5.	bɑ:sˁ bus	4		17.	dʊktɔ:r doctor	2		29.	ʃamʕa candle	2		41.	ħalazɔ:n snail	3	
6.	beɪjbI: baby	2		18.	jenɪtˁ jump	2		30.	jebu:s kiss	4		42.	teʃu:f see	2	
7.	festɑ:n dress	1		19.	ʔsˁbaʕ finger	4		31.	sˁeffɑ:ra whistle	1		43.	teχredʒ leave	3	
8.	ʕadʒɪ:na dough	3		20.	jeħfer dig	3		32.	teχɑ:nɪg scold	3		44.	jeχarreb destroy	4	
9.	bagara cow	4		21.	warda flower	4		33.	θʊʕbɑ:n snake	3		45.	tˁabla drum	1	
10.	bɐɪt house	1		22.	jeblaʕ gulp	3		34.	makɪ:nat χejɑ:tˁa sewing Machine	1		46.	mesmɑ:r nail	1	
11.	jIʃrab drink	4		23.	daʔɪra circle	3		35.	tˁɑ:wu:s peacock	2		47.	tegtˁef pick	4	
12.	bʊrtʊqɑ:l orange	2		24.	manfu:χ blown	1		36.	muʔddab polite	4		48.	ħemɑ:r donkey	3	
No. of errors				No. of errors				No. of errors				No. of errors			

Group 5				Group 6				Group 7			
49.	jeɡɪːs measure	2		61.	ʤɔːz ʔɪlhend coconut	2		73.	qalʕah castle	1	
50.	mekaʃʃer frawning	2		62.	jetsllaɡ climb	1		74.	mʊmarreðˤah nurse	2	
51.	ɡetɑːr guitar	4		63.	teleskɔːb telescope	4		75.	taʕɑːwn cooperation	2	
52.	kanʁar kangaroo	1		64.	sˤajjɑːd hunter	2		76.	ħaʃɪːʃ grass	4	
53.	jeɡɑːbɪl meet	3		65.	faχuːr proud	3		77.	jefattɪʃ search	3	
54.	χarbɑːn destroyed	1		66.	tefħasˤ examine	1		78.	ʔħfɑːd grandchildren	1	
55.	tasˤɑːdʊm collision	2		67.	qamħ wheat	2		79.	mʊhandIs engineer	3	
56.	rʊkbah knee	4		68.	bʊrʤ tower	4		80.	bɑːzella peas	4	
57.	ʕeʃ nest	3		69.	faras ɪlbħr seahorse	3		81.	χajɑːlɪː fictional	4	
58.	meʃawwek thorny	1		70.	kaʕab heel	4		82.	ɡʊfl lock	3	
59.	jerfes kick	3		71.	mʊftarɪs predator	1		83.	hʊdhʊd hoopoe	2	
60.	dɛɪnasˤɔːr dinosaur	4		72.	ðɛɪl tail	3		84.	ʃewajjah few	1	
No. of errors				No. of errors				No. of errors			

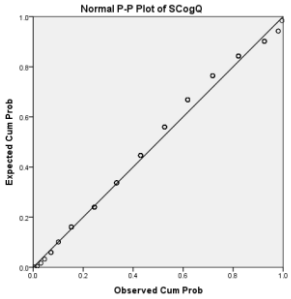
No. of ceiling item	
Minus total errors	-
Raw Score	=

APPENDIX H: P-P Plots for all measures, box plots and normality tests

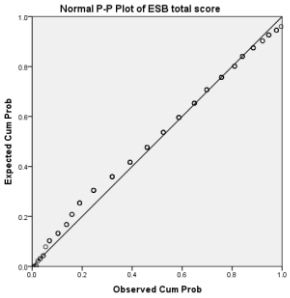
Arabic Research Adaptation of the LUI



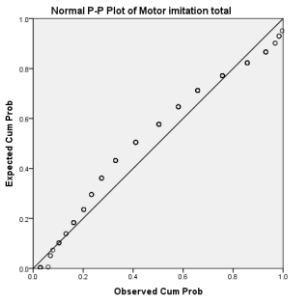
Sociocognitive Questionnaire



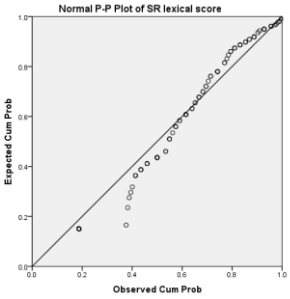
Early Sociocognitive Battery



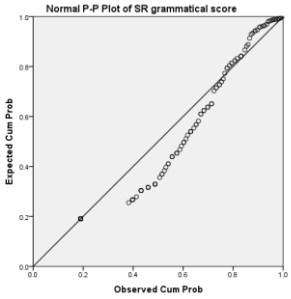
Motor Imitation



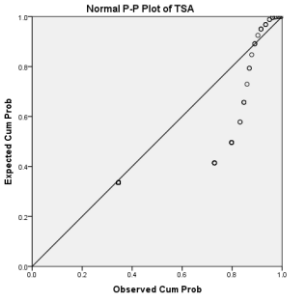
Sentence Repetition Lexical Morpheme Score



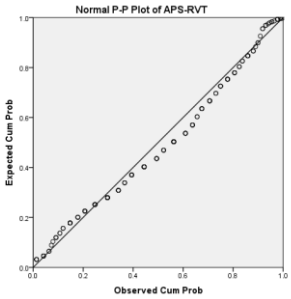
Sentence Repetition Grammatical Morpheme Score



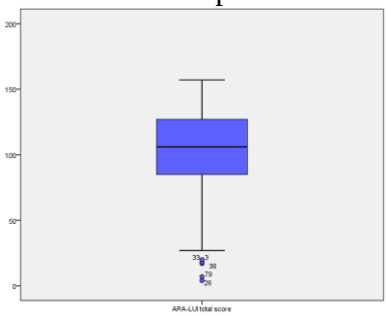
Total Sentence Accuracy



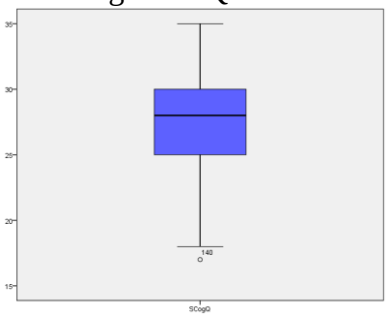
Arabic Preschool Receptive Vocabulary Test



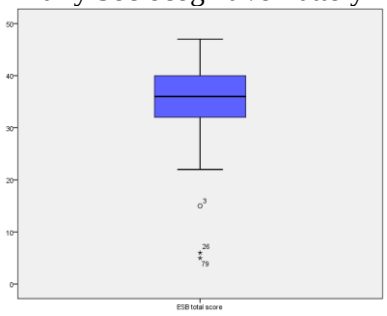
Arabic Research Adaptation of the LUI



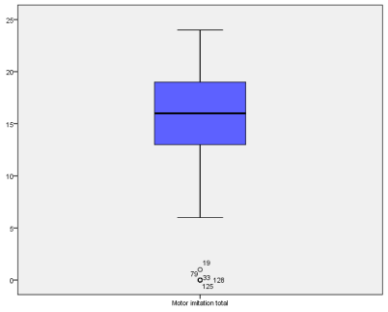
Sociocognitive Questionnaire



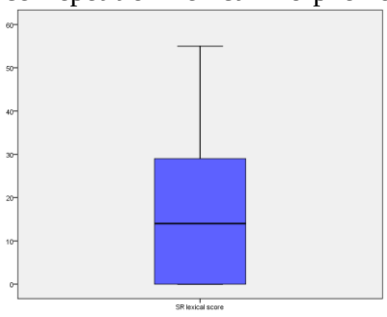
Early Sociocognitive Battery



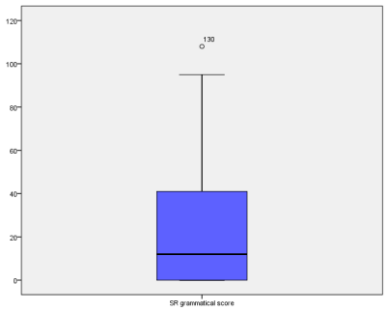
Motor Imitation



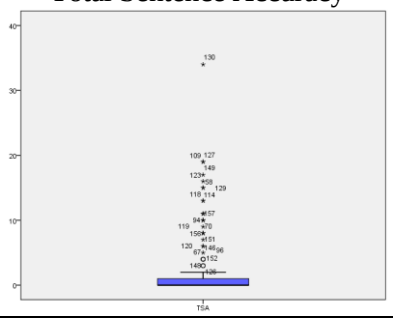
Sentence Repetition Lexical Morpheme Score



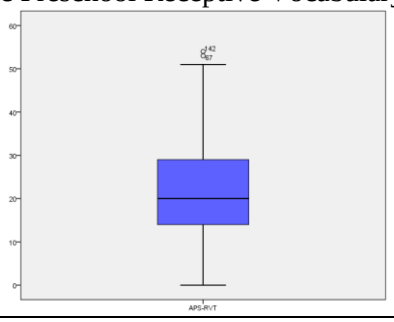
Sentence Repetition Grammatical Morpheme Score



Total Sentence Accuracy



Arabic Preschool Receptive Vocabulary Test



Normality Tests

ARA-LUI Shapiro-Wilk			
Age group	Statistic	df	Sig.
1	.918	46	.003
2	.903	46	.001
3	.926	42	.010

SCogQ Shapiro-Wilk			
Age group	Statistic	df	Sig.
1	.946	45	.035
2	.976	47	.439
3	.883	43	.000

ESB Kolmogorov-Smirnov			
Age group	Statistic	df	Sig.
1	.184	51	.000
2	.139	55	.010
3	.150	55	.003

APS-RVT Kolmogorov-Smirnov			
Age group	Statistic	df	Sig.
1	.129	51	.033
2	.122	55	.040
3	.201	55	.000

SR Lexical Kolmogorov-Smirnov			
Age group	Statistic	df	Sig.
2	.075	55	.200
3	.093	55	.200

SR Grammatical Kolmogorov-Smirnov			
Age group	Statistic	df	Sig.
2	.103	55	.200
3	.095	55	.200

TSA Kolmogorov-Smirnov			
Age group	Statistic	df	Sig.
2	.392	55	.000
3	.270	55	.000

APS-RVT Kolmogorov-Smirnov			
Age group	Statistic	df	Sig.
1	.088	51	.200
2	.149	55	.004
3	.126	54	.031

APPENDIX I: Non –Parametric analyses investigating effect of age in all assessment measures

Results of the Kruskal-Wallis test investigating effect of age on performance on all assessment measures

Test Name	Test Statistic (H)	df	sig
ARA-LUI	40.64	2	<.001
SCogQ	3.64	2	.162
ESB	54.38	2	<.001
Motor Imitation	39.24	2	<.001
APS-RVT	60.97	2	<.001

Significant results were followed up using separate Mann-Whitney for each pair of the age groups (i.e. groups 1 and 2, groups 2 and 3, groups 1 and 3) and the significance level was adjusted for multiple comparisons to $p < .016$

Results of Mann-Whitney U Test Comparing Performance of the Three Age Groups on Different Measures

Comparison of Performance of Age Groups on the Adapted LUI	U	z	P	r
age group 1-age group 2	545.50	-4.00	.000	-.42
age group 2-age group 3	545.00	-3.44	.001	-.37
age group 1- age group 3	269.50	-5.82	.000	-.62

Comparison of Performance of Age Groups on the ESB	U	z	P	r
age group 1-age group 2	794.00	-3.86	.000	-.37
age group 2-age group 3	716.00	-4.77	.000	-.45
age group 1- age group 3	343.00	-6.71	.000	-.65

Comparison of Performance of Age Groups on the Motor Imitation	U	z	P	r
age group 1-age group 2	866.50	-3.40	.001	-.33
age group 2-age group 3	910.00	-3.62	.000	-.34
age group 1- age group 3	473.50	-5.89	.000	-.57

Comparison of Performance of Age Groups on the APS-RVT	U	z	P	r
age group 1-age group 2	575.50	-5.24	.000	-.51
age group 2-age group 3	884.50	-3.64	.000	-.35
age group 1- age group 3	252.50	-7.21	.000	-.70

Results of Mann-Whitney U Test Comparing Performance between age group 1 and 2 on Sentence Repetition Test

Comparison of Performance of Age Groups on the Sentence Repetition	**U**	**z**	**P**	**r**
SR Lexical age group 2-age group 3	1039.50	-2.83	.005	-.27
SR Grammatical age group 2-age group 3	1033.50	-2.87	.004	-.27
TSA age group 2- age group 3	926.00	-3.84	.000	-.37

APPENDIX J: Comparison of scores of children on the ARA-LUI and SR in the current study to previous research

Comparison of the scores of the LUI (O'Neill, 2009) and ARA-LUI

Age group	LUI[a]	ARA-LUI[b]
	(Maximum score = 161)	
1 (2;0-2;5)	93	84
2 (2;6-2;11)	125.75	109
3 (3;0-3;5)	142.75	128

[a] 50th percentile score calculated from the corresponding age range for girls and boys

[b] median score

Comparison of the Sentence Repetition scores in Wallan, Chiat and Roy's (2011) study and the current study

Age group	n		SR Lexical maximum score = 56 mean (SD)		SR Grammatical maximum score = 117 mean (SD)		TSA maximum score = 42 mean (SD)	
2 (2;6-2;11)	20	55	21.85 (9.39)	20.53 (11.70)	29.95 (21.67)	27.16 (20.30)	1.25 (2.45)	1.29 (3.04)
3 (3;0-3;5)	20	55	30.20 (10.55)	28.07 (14.74)	50.90 (25.73)	43.84 (30.15)	4.9 (7.77)	4.71 (6.94)

Scores in red are for the current study, scores in black are for Wallan and colleagues' study (A. Wallan, personal communication, October, 25, 2014)

APPENDIX K: Parental concern in relation to children's performance on the different measures

Parental concern	Performance on Test Battery
No concern 114 (70.80%)	78 average (68.42%) 24 low on one test: • 19 sociocognitive (16.67%) • 5 language (4.39%) 2 Profile 3 - language (1.75%) 9 Profile 4 - mixed (7.89%) 1 Profile 1 - broad delay (0.87%)
Suspected 16 (9.94%)	6 average (37.5%) 2 low on one test: • 2 language (12.5%) 2 Profile 3 - language (12.5%) 5 Profile 4 - mixed (31.25%) 1 Profile 1- broad delay (6.25%)
Diagnosed 4 (2.48%)	2 average (50%) 1 low on one language test (25%) 1 Profile 3- language (25%)
Missing information 27 (16.77%)	24 incomplete data (88.89%) 1 Profile 2 - sociocognitive (3.7%) 1 Profile - mixed (3.7%) 1 Profile 1 - broad delay (3.7%)

REFERENCES

Abravanel, E., Levan-Goldschmidt, E., & Stevenson, M. B. (1976). Action imitation: The early phase of infancy. *Child Development, 47*(2), 1032-1044.

Abu Allam, R., & Hadi, F. (1998). *Picture Vocabulary Test*. Kuwait: CCET.

Akechi, H., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2013). Brief report: Pointing cues facilitate word learning in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(1), 230-235.

Akechi, H., Senju, A., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2011). Do children with ASD use referential gaze to learn the name of an object? an eye-tracking study. *Research in Autism Spectrum Disorders, 5*(3), 1230-1242.

Alkadhi, A. (2010). *The performance of 2-4 year old children in Saudi Arabia on three sociocognitive tasks*. Unpublished MSc Dissertation, City University London, London.

American Speech-Language-Hearing Association. (2008). *Roles and responsibilities of speech-language pathologists in early intervention: Guidelines*. Available from www.asha.org/policy.

Bacon, A. L., Fein, D., Morris, R., Waterhouse, L., & Allen, D. (1998). The responses of autistic children to the distress of others. *Journal of Autism and Developmental Disorders, 28*(2), 129-142.

Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., & Drew, A. (2000). A screening instrument for autism at 18 months of age: A 6-year follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(6), 694-702.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*(02), 395-418.

Baranek, G., Watson, L., Crais, E., & Reznick, J. (2003). *First Year Inventory (FYI)(version 2.0)*. Chapel Hill, NC: The University of North Carolina at Chapel Hill.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development, 68*(1), 48-57.

Bates, E. (1979). Intentions, conventions, and symbols. In E. Bates, L. Benigni, I. Bretherton, L. Camaioni, & V. Volterra (Eds.), *The emergence of symbols: Cognition and communication in infancy* (pp. 33-68). New York: Academic Press.

Bates, E., Bretherton, I., Snyder, L., Shore, C., & Volterra, V. (1980). Vocal and gestural symbols at 13 months. *Merrill-Palmer Quarterly of Behavior and Development, 26*(4), 407-423.

Bates, E., Thal, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology, 25*(6), 1004- 1019.

Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio, TX: Harcourt Assessment Inc.

Beadle-Brown, J. D., & Whiten, A. (2004). Elicited imitation in children and adults with autism: Is there a deficit? *Journal of Intellectual and Developmental Disability, 29*(2), 147-163.

Begeer, S., Koot, H. M., Rieffe, C., Terwogt, M., & Stegge, H. (2007). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review, 28*(3), 342-369.

Berman, J. M., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology, 107*(2), 87-99.

Berman, J. M., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child Development, 84*(5), 1791-1805.

Beuker, K. T., Rommelse, N. N., Donders, R., & Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behavior and Development, 36*(1), 71-83.

Bigham, S. (2008). Comprehension of pretence in children with autism. *British Journal of Developmental Psychology, 26*(2), 265-280.

Bishop, D. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders, 49*(4), 381-415.

Bishop, D. V. (2003). *The Children's Communication Checklist Version 2* (CCC-2). London: Psychological Corporation.

Bishop, D. V., & Norbury, C. F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: A study using standardised diagnostic instruments. *Journal of Child Psychology and Psychiatry, 43*(7), 917-929.

Bishop, D. V., Holt, G., Line, E., McDonald, D., McDonald, S., & Watt, H. (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to 4 years: A longitudinal study of precursors of specific language impairment. *Journal of Neurodevelopmental Disorders, 4*(1), 1-12.

Bishop, D. V., Price, T. S., Dale, P. S., & Plomin, R. (2003). Outcomes of early language DelayII. etiology of transient and persistent language difficulties. *Journal of Speech, Language, and Hearing Research, 46*(3), 561-575.

Bornstein, M. H., Hahn, C., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24*(3), 267-304.

Botting, N. (2005). Non-verbal cognitive development and language impairment. *Journal of Child Psychology and Psychiatry, 46*(3), 317-326.

Boucher, J., & Lewis, V. (1997). *Pre-school Language Scale-3* (UK Adaptation). London: Harcourt Brace & Company Books.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*(6), 535-543.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language, 35*(01), 207-220.

Brooks, R., & Meltzoff, A. N. (2014). Gaze following: A mechanism for building social connections between infants and adults. In M. Mikulincer & P. R. Shaver (Eds.),

Mechanisms of social connection: From brain to group (pp. 167-183). Washington, DC: American Psychological Association.

Broomfield, J., & Dodd, B. (2011). Is speech and language therapy effective for children with primary speech and language impairment? report of a randomized control trial. *International Journal of Language & Communication Disorders, 46*(6), 628-640.

Buschmann, A., Jooss, B., Rupp, A., Dockter, S., Blaschtikowitz, H., Heggen, I., & Pietz, J. (2008). Children with developmental language delay at 24 months of age: Results of a diagnostic work-up. *Developmental Medicine & Child Neurology, 50*(3), 223-229.

Butterworth, G. (1999). Neonatal imitation: Existence, mechanisms and motives. In J. Nadel & G. Butterworth (Eds.), *Imitation in Infancy* (pp. 63–88), Cambridge: Cambridge University Press.

Capone, N. C., & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language and Hearing Research, 47*(1), 173-186.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seeman (Ed.), *Joint Attention: New Developments in Psychology, Philosphy of Mind, and Social Neuroscience* (pp. 159-182). Cambridge: MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i-174.

Carpenter, M., Pennington, B. F., & Rogers, S. J. (2002). Interrelations among social-cognitive skills in young children with autism. *Journal of Autism and Developmental Disorders, 32*(2), 91-106.

Casby, M. W. (1997). Symbolic play of children with language Impairment. A critical review. *Journal of Speech, Language, and Hearing Research, 40*(3), 468-479.

Charman, T., & Baron-Cohen, S. (1994). Another look at imitation in autism. *Development and Psychopathology, 6*(3), 403-413.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development, 15*(4), 481-498.

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology, 33*(5), 781-789.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions, and evidence. *Language and Cognitive Processes in Developmental Disorders, 16*(2/3), 113-142.

Chiat, S., & Roy, P. (2006a). Very early processing skills as predictors of later language disorders. Retrieved August 28, 2009, from http://www.esrcsocietytoday.ac.uk/.

Chiat, S., & Roy, P. (2006b). Early Sociocognitive Battery. Retrieved November 15, 2009, from http://www.city.ac.uk/lcs/research/veps/assess.html.

Chiat, S., & Roy, P. (2007). The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research, 50*(2), 429-443.

Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry, 49*(6), 635-645.

Chiat, S., & Roy, P. (2013). Early predictors of language and social communication impairments at 9-11 years: A follow-up study of early referred children. *Journal of Speech, Language and Hearing Research, 56,* 1824-1836.

Clifford, S., & Dissanayake, C. (2009). Dyadic and triadic behaviours in infancy as precursors to later social responsiveness in young children with autistic disorder. *Journal of Autism and Developmental Disorders, 39*(10), 1369-1380.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Constantino, J. N., & Gruber, C. P. (2005). *Social Responsiveness Scale*. Torrance, CA: Western Psychological Services.

Cripe, J. J. W., & Slentz, K. (1993). *Assessment, evaluation, and programming system for infants and children: AEPS measurement for birth to three years.* Baltimore, MD: Paul H Brookes Pub Co.

Dahlgren, S., & Sandberg, A. D. (2008). Referential communication in children with autism spectrum disorder. *Autism : The International Journal of Research and Practice, 12*(4), 335-348.

Dale, P. S. (1996). Parent report assessment of language and communication. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (pp. 161–182). Baltimore: Paul H. Brookes.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46*(3), 544-560.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*(2), 271-283.

Decety, J. (2006). A cognitive neuroscience view of imitation. In S. J. Rogers & J. H. G. Williams (Eds.), *Imitation and the Social Mind: Autism and Typical Development* (pp. 251-276), New York, London: Guilford Press.

Delgado, C. E., Mundy, P., Crowson, M., Markus, J., Yale, M., & Schwartz, H. (2002). Responding to joint attention and language DevelopmentA comparison of target locations. *Journal of Speech, Language, and Hearing Research, 45*(4), 715-719.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science, 238*(4833), 1556-1557.

DeLoache, J. S. (1989). Young children understanding of the correspondence between a scale model and a larger space. *Cognitive Development, 4*(2), 121-139.

DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development, 62*(4), 736-752.

DeLoache, J. S. (2002). The symbol-mindedness of young children. *Child Psychology in Retrospect and Prospect: In Celebration of the 75th Anniversary of the Institute of Child Development, 32*, 73-101.

DeMyer, M. K., Alpern, G. D., Barton, S., DeMyer, W. E., Churchill, D. W., Hingtgen, J. N., . . . Kimberlin, C. (1972). Imitation in autistic, early schizophrenic, and non-psychotic subnormal children. *Journal of Autism and Childhood Schizophrenia, 2*(3), 264-287.

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2010). Three profiles of language abilities in toddlers with an expressive vocabulary delay: Variations on a theme. *Journal of Speech, Language and Hearing Research, 53*(3), 699-709.

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43*(4), 361-389.

DeVellis, R. F. (2012). *Scale development: Theory and applications.* Los Angeles: Sage Publications.

DiLavore, P., Lord, C., & Rutter, M. (1995). The Prelinguistic autism diagnostic observation schedule. *Journal of Autism and Developmental Disorders, 25*, 355–379.

Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health, 20*(2), 116-125.

Dohmen, A. (2010). *An investigation of nonverbal imitation and language in children with specific language delay.* Unpublished doctoral dissertation, City University London, London.

Dohmen, A., Chiat, S., & Roy, P. (2013). Nonverbal imitation skills in children with specific language delay. *Research in Developmental Disabilities, 34*(10), 3288-3300.

Dollaghan, C. A. (2004). Evidence-based practice in communication disorders: What do we know, and when do we know it? *Journal of Communication Disorders, 37*(5), 391-400.

Doswell, G., Lewis, V., Sylva, K., & Boucher, J. (1994). Validational data on the warwick symbolic play test. *International Journal of Language & Communication Disorders, 29*(3), 289-298.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test: 4th edition.* San Antonio, TX: Pearson.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale*(2nd ed.) . Windsor: NFER-Nelson publishing company ltd.

Dunn, L.M., Dunn, D.M., & Styles, B. (2009). *British Picture Vocabulary Scale: 3rd edition.* London: GL assessment.

Eadie, P. A., Ukoumunne, O., Skeat, J., Prior, M. R., Bavin, E., Bretherton, L., & Reilly, S. (2010). Assessing early communication behaviours: Structure and validity of the communication and symbolic behaviour Scales—Developmental profile (CSBS-DP) in 12-month-old infants. *International Journal of Language & Communication Disorders, 45*(5), 572-585.

Eadie, P., Reilly, S., Bavin., E., Bretherton, L., Carlin, J., Castles, A., … Duns, R. (2013). *The association between early communication & symbolic behaviours of infants at 12 months and diagnosis of language impairment or autism at 4 years.* Paper presented at the child language seminars, Manchester, UK.

Elliot, C. (1990). *Manual for the Differential Abilities Scales*. San Antonio, TX: Psychological Corporation.

Ellis, E. M., & Thal, D. J. (2008). Early language delay and risk for language impairment. *SIG 1 Perspectives on Language Learning and Education, 15*(3), 93-100.

Eriksson, M., Westerlund, M., & Berglund, E. (2002). A screening version of the Swedish communicative development inventories designed for use with 18-month-old children. *Journal of Speech, Language and Hearing Research, 45*(5), 948-960.

Everitt, A., Hannaford, P., & Conti-Ramsden, G. (2013). Markers for persistent specific expressive language delay in 3–4-year-olds. *International Journal of Language & Communication Disorders, 48*(5), 534-553.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur communicative

development inventories at ages one and two years. *Child Development, 71*(2), 310-322.

Fensen, L., Dale, P.,Reznick, S., Thal, D., Bates, E., et al. (2002). *MacArthur Communicative Developmental Inventories.* 2nd ed. Baltimore: Paul H. Brookes.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *The MacArthur-Bates Communicative Development Inventories user's guide and technical manual, 2nd ed. Baltimore, MD: Brookes.*

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Reply: Measuring variability in early child language: Don't shoot the messenger. *Child Development, 71*(2), 323-328.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i-185.

Fenson, L., Dale, P. S., Reznick, J. S., Thai, D., Mates, E., Hartung, J. P., et al. (1993). *MacArthur Communicative Developmental Inventories: User's guide and technical manual.* San Diego: Singular.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics 4th Ed.* London: Sage Publications Ltd.

Fujiki, M., Spackman, M. P., Brinton, B., & Illig, T. (2008). Ability of children with language impairment to understand emotion conveyed by prosody in a narrative passage. *International Journal of Language & Communication Disorders, 43*(3), 330-345.

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 57*(4), 1363-1382.

Gatt, D., Grech, H., & Dodd, B. (2013). Early lexical expression in typically developing Maltese children: Implications for the identification of language delay. *Clinical Linguistics & Phonetics, 27*(6-7), 459-471.

Ghuman, J. K., Freund, L., Reiss, A., Serwint, J., & Folstein, S. (1998). Early detection of social interaction problems: Development of a social interaction instrument in young children. *Journal of Developmental and Behavioral Pediatrics, 19*(6), 411-419.

Gilliam, J. E. (1995). *Gilliam autism rating scale*. Austin, TX: Pro-Ed.

Girolametto, L., Wiigs, M., Smyth, R., Weitzman, E., & Pearce, P. S. (2001). Children with a history of expressive vocabulary delay: Outcomes at 5 years of age. *American Journal of Speech-Language Pathology, 10*(4), 358-369.

Gleissner, B., Bekkering, H., & Meltzoff, A. N. (2000). Children's coding of human action: Cognitive factors influencing imitation in 3-year-old. *Developmental Science, 3*(4), 405-414.

Goodman, R. (1997). *Strengths and Difficulties Questionnaire*. Retrieved January, 2010, from www.sdqinfo.com

Goodwyn, S. W., Acredolo, L. P., & Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior, 24*(2), 81-103.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory. *Journal of Child Language, 27*(3), 689-705.

Hart, K. I., Fujiki, M., Brinton, B., & Hart, C. H. (2004). The relationship between social behavior and severity of language impairment. *Journal of Speech, Language and Hearing Research, 47*(3), 647-662.

Hedrick, D. L., Prather, E. M., & Tobin, A. R. (1975). *Sequenced Inventory of Communication Development.* Seattle: University of Washington Press.

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W., Hofman, A., . . . Tiemeier, H. (2011). Examining continuity of early expressive vocabulary development: The generation R study. *Journal of Speech, Language and Hearing Research, 54*(3), 854-869.

Hill, E. L. (1998). A dyspraxic deficit in specific language impairment and developmental coordination disorder? evidence from hand and arm movements. *Developmental Medicine & Child Neurology, 40*(6), 388-395.

Horwitz, S., Irwin, J. R., Briggs-Gowan, M. J., Bosson Heenan, J. M., Mendoza, J., & Carter, A. S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(8), 932-940.

Hutman, T., Rozga, A., DeLaurentis, A. D., Barnwell, J. M., Sugar, C. A., & Sigman, M. (2010). Response to distress in infants at risk for autism: A prospective longitudinal study. *Journal of Child Psychology and Psychiatry, 51*(9), 1010-1020.

Hutman, T., Rozga, A., DeLaurentis, A., Sigman, M., & Dapretto, M. (2012). Infants' pre-empathic behaviors are associated with language skills. *Infant Behavior and Development, 35*(3), 561-569.

Ingersoll, B. (2012). Brief report: Effect of a focused imitation intervention on social functioning in children with autism. *Journal of Autism and Developmental Disorders,42*(8), 1768-1773.

Ingersoll, B., & Meyer, K. (2011). Do object and gesture imitation skills represent independent dimensions in autism? *Journal of Developmental and Physical Disabilities, 23*(5), 421-431.

Ingersoll, B., & Schreibman, L. (2006). Teaching reciprocal imitation skills to young children with autism using a naturalistic behavioral approach: Effects on language, pretend play, and joint attention. *Journal of Autism and Developmental Disorders, 36*(4), 487-505.

Ingersoll, B., Lewis, E., & Kroman, E. (2007). Teaching the imitation and spontaneous use of descriptive gestures in young children with autism using a naturalistic behavioral intervention. *Journal of Autism and Developmental Disorders, 37*(8), 1446-1456.

Jansen, R., Ceulemans, E., Grauwels, J., Maljaars, J., Zink, I., Steyaert, J., & Noens, I. (2013). Young children with language difficulties: A dimensional approach to subgrouping. *Research in Developmental Disabilities, 34*(11), 4115-4124.

Keegstra, A., Knijff, W., Post, W., & Goorhuis-Brouwer, S. (2007). Children with language problems in a speech and hearing clinic: Background variables and extent of language problems. *International Journal of Pediatric Otorhinolaryngology, 71*(5), 815-821.

Ketelaars, M. P., Cuperus, J., Jansonius, K., & Verhoeven, L. (2010). Pragmatic language impairment and associated behavioural problems. *International Journal of Language & Communication Disorders, 45*(2), 204-214.

Klee, T., Pearce, K., & Carson, D. K. (2000). Improving the positive predictive value of screening for developmental language disorder. *Journal of Speech, Language, and Hearing Research, 43*(4), 821-833.

Laakso, M. L., Poikkeus, A. M., Eklund, K., & Lyytinen, P. (1999). Social interactional behaviors and symbolic play competence as predictors of language development and their associations with maternal attention-directing strategies. *Infant Behavior and Development, 22*(4), 541-556.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological Review, 94*(4), 412-426.

Lewis, V., & Boucher, J. (1997). *Manual of the Test of Pretend Play*. London: The Psychological Corporation.

Lewis, V., Boucher, J., Lupton, L., Watson, S. (2000). Relationships between symbolic play, functional play, verbal and non-verbal ability in young children. *International Journal of Language & Communication Disorders, 35*(1), 117-127.

Leyfer, O. T., Tager-Flusberg, H., Dowd, M., Tomblin, J. B., & Folstein, S. E. (2008). Overlap between autism and specific language impairment: Comparison of autism diagnostic interview and autism diagnostic observation schedule scores. *Autism Research, 1*(5), 284-296.

Lillard, A. (2007). Guided participation: How mothers structure and children understand pretend play. In A. Göncü & S. Gaskins (Eds.), *Play and development: Evolutionary, sociocultural, and functional perspectives* (pp. 131-154). Mahwah, NJ: Lawrence Erlbaum Associates.

Lillard, A. S., Pinkham, A., & Smith, E. (2011). Pretend play and cognitive development. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 285-311). West Sussex, UK: Blackwell Publishing Ltd.

Lillard, A., & Kavanaugh, R. (2014). A longitudinal study of the relationship between symbolic understanding and theory of mind in preschoolers. *Child Development,* (In press).

Lin, C., & Chiu, C. (2014). Adaptation of the chinese edition of the CSBS-DP: A cross-cultural comparison of prelinguistic development between taiwanese and american toddlers. *Research in Developmental Disabilities, 35*(5), 1042-1050.

Lord, C., Risi, S., Lambrecht, L., Cook Jr, E. H., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The autism diagnostic observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(8), 1426-1438.

Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology, 45*(6), 1774-1786.

MacConnell, A., & Daehler, M. W. (2004). The development of representational insight: Beyond the model/room paradigm. *Cognitive Development, 19*(3), 345-362.

Maljaars, J., Noens, I., Scholte, E., & van Berckelaer-Onnes, I. (2012). Language in low-functioning children with autistic disorder: Differences between receptive and expressive skills and concurrent predictors of language. *Journal of Autism and Developmental Disorders, 42*(10), 2181-2191.

Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with asperger syndrome. *Journal of Autism and Developmental Disorders, 34*(3), 311-328.

Marton, K. (2009). Imitation of body postures and hand movements in children with specific language impairment. *Journal of Experimental Child Psychology, 102*(1), 1-13.

Mayers, A. (2013). *Introduction to statistics and SPSS in psychology.* London: Pearson.

McCathren, R. B., Yoder, P. J., & Warren, S. F. (2000). Testing predictive validity of the communication composite of the communication and symbolic behavior scales. *Journal of Early Intervention, 23*(1), 36-46.

McCune, L. (1995). A normative study of representational play in the transition to language. *Developmental Psychology, 31*(2), 198.

McDuffie, A., Yoder, P., & Stone, W. (2005). Prelinguistic predictors of vocabulary in young children with autism spectrum disorders. *Journal of Speech, Language and Hearing Research, 48*(5), 1080-1097.

McEwen, F., Happe, F., Bolton, P., Rijsdijk, F., Ronald, A., Dworzynski, K., & Plomin, R. (2007). Origins of individual differences in imitation: Links with language, pretend play, and socially insightful behavior in Two-Year-Old twins. *Child Development, 78*(2), 474-492.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology, 24*(4), 470-476.

Meltzoff, A. N. (2005). Imitation and other minds: The "like me" hypothesis. In S. Hurley, & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science-volume 2: Imitation, human development, & culture* (pp. 55-77). Massachusetts: MIT Press.

Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Developmental Science, 10*(1), 126-134.

Meltzoff, A. N. (2011). Social cognition and the origins of imitation, empathy, and theory of mind. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed.) (pp. 49-75). Oxford: Willey-Blackwell.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 75-78.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development, 54*(3), 702-709.

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology, 25*(6), 954-962.

Miniscalco, C., Rudling, M., Råstam, M., Gillberg, C., & Johnels, J. Å. (2014). Imitation (rather than core language) predicts pragmatic development in young children with ASD: A preliminary longitudinal study using CDI parental reports. *International Journal of Language & Communication Disorders, 49*(3), 369-375.

Montague, D., & Walker-Andrews, A. (2001). Peakaboo: A new look at infants' perception of emotion expression. *Developmental Psychology, 37*(6), 826-838.

Morales, M., Mundy, P., Delgado, C. E., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6-through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology, 21*(3), 283-298.

Mullen, E. M. (1985). *Manual for the Infant Mullen Scales of Early Learning*. Cranston, RI .

Mullen, E. M. (1989). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service, Inc.

Mullen, E. M. (1995). *Mullen Scales of Early Learning*: AGS edition. Circle Pines: American Guidance Service.

Mullen, E. M. (1997). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service, Inc.

Mundy, P., & Gomes, A. (1997). A skills approach to early language development: Lessons from research on developmental disabilities. In L. B. Adamson & M. A. Romski (Eds.), *Communication and language acquisition: Discoveries from atypical development* (pp. 107-133). Baltimore: Brookes.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*(3), 938-954.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *A manual for the abridged Early Social Communication Scales (ESCS).* (Unpublished manuscript) Coral Gables, FL: Department of Psychology, University of Miami.

Mundy, P., Hogan, A., & Doehring, P. (1996). *A preliminary manual for the abridged Early Social-Communication Scales.* (Unpublished Manuscript) Coral Gables, FL: University of Miami.

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry*, *27*(5), 657-669.

Nadel, J. (2002). Imitation and imitation recognition: Functional use in preverbal infants and nonverbal children with autism. In A. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain Bases* (pp. 42-62). Cambridge: Cambridge University Press.

Nadel, J. (2006). Does imitation matter to children with autism. In S.J. Rogers & J. H. Williams (Eds.), *Imitation and the development of the social mind: Lessons from typical development and autism* (pp. 118-137). New York: Guilford Press.

Nadel, J., Guérini, C., Pezé, A., & Rivet, C. (1999). The evolving nature of imitation as a format for communication. . In J. Nadel & G. Butterworth (Eds.), *Imitation in Infancy* (pp. 63–88), Cambridge: Cambridge University Press.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*(3), 555-565.

O'Neill, D. (2009). *Language Use Inventory: An assessment for young children's pragmatic language*. Waterloo, Canada: Knowledge in Development.

O'Neill, D. K. (2007). The language use inventory for young children: A parent-report measure of pragmatic language development for 18-to 47-month-old children. *Journal of Speech, Language, and Hearing Research*, *50*(1), 214-228.

O'Neill, D. K. (2014). Assessing pragmatic language functioning in young children. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition* (pp. 363-386). Amesterdam: John Benjamins Publishing.

O'Neill, D. K., & Atance, C. M. (2000). 'Maybe my daddy give me a big piano': The development of children's use of modals to express uncertainty. *First Language, 20*(58), 029-52.

O'Neill, H., & Chiat, S. (2015). What our hands say: Exploring gesture use in subgroups of children with language delay. *Journal of Speech, Language, and Hearing Research, 58*(4), 1319-1325.

O'toole, C., & Chiat, S. (2006). Symbolic functioning and language development in children with down syndrome. *International Journal of Language & Communication Disorders, 41*(2), 155-171.

Page, J., & Boucher, J. (1998). Motor impairments in children with autistic disorder. *Child Language Teaching and Therapy, 14*(3), 233-259.

Pan, B., Rowe, M. L., Spier, E., & Tamis-Lemonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language, 31*(03), 587-608.

Paul, R., & Roth, F. P. (2011). Characterizing and predicting outcomes of communication delays in infants and toddlers: Implications for clinical practice. *Language, Speech, and Hearing Services in Schools, 42*(3), 331-340.

Paul, R., & Shiffer, M. E. (1991). Communicative initiations in normal and late-talking toddlers. *Applied Psycholinguistics, 12*(04), 419-431.

Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language outcomes of toddlers with autism spectrum disorders: A two year follow-up. *Autism Research, 1*(2), 97-107.

Paul, R., Looney, S. S., & Dahm, P. S. (1991). Communication and socialization skills at ages 2 and 3 in" late-talking" young children. *Journal of Speech, Language and Hearing Research, 34*(4), 858-865.

Pena, E. D., Spaulding, T. J., & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology, 15*(3), 247-254.

Pesco, D., & O'Neill, D. K. (2012). Predicting later language outcomes from the language use inventory. *Journal of Speech, Language and Hearing Research, 55*(2), 421-434.

Piaget, J. (1962). *Play, dreams and imitation.* New York: Norton.

Pickard, K. E., & Ingersoll, B. R. (2015). Brief report: High and low level initiations of joint attention, and response to joint attention: Differential relationships with language and imitation. *Journal of Autism and Developmental Disorders, 45*(1), 262-268.

Poll, G. H., & Miller, C. A. (2013). Late talking, typical talking, and weak language skills at middle childhood. *Learning and Individual Differences, 26,* 177-184.

Poon, K. K., Watson, L. R., Baranek, G. T., & Poe, M. D. (2012). To what extent do joint attention, imitation, and object play behaviors in infancy predict later communication and intellectual functioning in ASD? *Journal of Autism and Developmental Disorders, 42*(6), 1064-1074.

Rakoczy, H., & Tomasello, M. (2006). Two-year-olds grasp the intentional structure of pretense acts. *Developmental Science, 9*(6), 557-564.

Rakoczy, H., Tomasello, M., & Striano, T. (2004). Young children know that trying is not pretending: A test of the" behaving-as-if" construal of children's early concept of pretense. *Developmental Psychology, 40*(3), 388-399.

Reddy, V., Williams, E., & Vaughan, A. (2002). Sharing humour and laughter in autism and down's syndrome. *British Journal of Psychology, 93*(2), 219-242.

Reilly, S., Bishop, D. V., & Tomblin, B. (2014). Terminological debate over language impairment in children: Forward movement and sticking points. *International Journal of Language & Communication Disorders, 49*(4), 452-462.

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., . . . Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from early language in Victoria study. *Pediatrics, 126*(6), e1530-e1537.

Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: Infants selectively respond to indirect emotional signals. *Child Development, 78*(2), 503-521.

Repacholi, B. M., Meltzoff, A. N., & Olsen, B. (2008). Infants' understanding of the link between visual perception and emotion:" if she can't see me doing it, she won't get angry.". *Developmental Psychology, 44*(2), 561-574.

Repacholi, B. M., Meltzoff, A. N., Rowe, H., & Toub, T. S. (2014). Infant, control thyself: Infants' integration of multiple social cues to regulate their imitative behavior. *Cognitive Development, 32*, 46-57.

Rescorla, L. (1989). The language development survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*(4), 587-599.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language and Hearing Research, 45*(2), 360-371.

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language and Hearing Research, 48*(2), 459-472.

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language and Hearing Research, 52*(1), 16-30.

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*(2), 141-150.

Rescorla, L., & Goossens, M. (1992). Symbolic play development in toddlers with expressive specific language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research, 35*(6), 1290-1302.

Rescorla, L., & Schwartz, E. (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics, 11*(04), 393-407.

Rescorla, L., Bascome, A., Lampard, J., & Feeny, N. (2001). Conversational patterns in late talkers at age 3. *Applied Psycholinguistics, 22*(02), 235-251.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3; 0 and 4; 0. *Journal of Child Language, 27*(3), 643-664.

Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: Lexical development from 2; 0 to 3; 0. *Journal of Child Language, 27*(02), 293-311.

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language and Hearing Research, 40*(3), 556-566.

Reynell, J. (1977). *Reynell Developmental Language Scales*. Windsor: NFER.

Reynell, JK.,Gruber, CP. (1990). *Reynell Developmental Language Scales*. Los Angeles: Western Psychological Services.

Rice, M. L., Sell, M. A., & Hadley, P. A. (1991). Social interactions of speech, and language-impaired children. *Journal of Speech, Language and Hearing Research, 34*(6), 1299-1307.

Ritvo, S., & Provence, S. (1953). Form perception and imitation in some autistic children: Diagnostic findings and their contextual interpretation. *Psychoanalytic Study of the Child, 8*, 155-161.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*(2), 131-144.

Rogers, S. J., & Pennington, B. F. (1991). A theoretical approach to the deficits in infantile autism. *Development and Psychopathology, 3*(2), 137-162.

Rogers, S. J., Hepburn, S. L., Stackhouse, T., & Wehner, E. (2003). Imitation performance in toddlers with autism and those with other developmental disorders. *Journal of Child Psychology and Psychiatry, 44*(5), 763-781.

Rossetti, L. M. (1990). *The Rossetti Infant-Toddler Language Scale: A measure of communication and interaction.* East Moline, IL: LinguiSystems.

Roy, P., & Chiat, S. (2014). Developmental pathways of language and social communication problems in 9–11 year olds: Unpicking the heterogeneity. *Research in Developmental Disabilities, 35*(10), 2534-2546.

Roy, P., Kersley, H. & Law, J. (2005). The Sure Start Measure Standardisation Study. Available from http://www.surestart.gov.uk/doc/P0001797.pdf.

Rust, J., & Golombok, S. (2009). *Modern psychometrics: The science of psychological assessment.* East Sussex: Psychology Press.

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

Sachse, S., & Von Suchodoletz, W. (2008). Early identification of language delay by direct language assessment or parent report? *Journal of Developmental and Behavioral Pediatrics, 29*(1), 34-41.

Sackett, D., Haynes, R. B., Guyatt, G., & Tugwell, P. (1991). *Clinical epidemiology: A basic Science for clinical medicine*. Boston: Little, Brown & Company.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*(01), 1-22.

Schaerlaekens, A., Zink, I., & Van Ommeslaeghe, K. (2003). *Reynell Taalontwikkelingsschalen. Handleiding-Tweedeversie.* Lisse: Swetz & Zeitlinger.

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language & Communication Disorders, 45*(6), 691-702.

Seeff-Gabriel, B., Chiat, S., & Roy, P. (2008). *Early Repetition Battery.* London: Pearson Assessment.

Shaalan, S. (2010). *Investigating grammatical complexity in Gulf Arabic speaking children with specific language impairment.* Unpublished doctoral dissertation, University College London, London.

Siebert, J., Hogan, A., & Mundy, P. (1982). Assessing international competencies: The early social communicative scales. *Infant Mental Health Journal*, 3, 137–162.

Sigman, M. D., Kasari, C., Kwon, J., & Yirmiya, N. (1992). Responses to the negative emotions of others by autistic, mentally retarded, and normal children. *Child Development, 63*(4), 796-807.

Sigman, M., & McGovern, C. W. (2005). Improvement in cognitive and language skills from preschool to adolescence in autism. *Journal of Autism and Developmental Disorders, 35*(1), 15-23.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64*(1), 1-114.

Sigman, M., & Ungerer, J. A. (1984). Cognitive and language skills in autistic, mentally retarded, and normal children. *Developmental Psychology, 20*(2), 293-302.

Siller, M., & Sigman, M. (2008). Modeling longitudinal change in the language abilities of children with autism: Parent behaviors and child characteristics as predictors of change. *Developmental Psychology, 44*(6), 1691-1704.

Spackman, M. P., Fujiki, M., & Brinton, B. (2006). Understanding emotions in context: The effects of language impairment on children's ability to infer emotional reactions. *International Journal of Language & Communication Disorders, 41*(2), 173-188.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales: Interview edition, survey form manual.* Circle Pines, MN: American Guidance Services.

Spencer, P. E. (1996). The association between language and symbolic play at two years: Evidence from deaf toddlers. *Child Development, 67*(3), 867-876.

Squires, J., Bricker, D., & Twombly, E. (2002). *Ages and Stages Questionnaire: Social-Emotional (ASQ: SE).* Baltimore, MD: Paul H.

St. James PJ, Tager-Flusberg H. (1994). An observational study of humor in autism and Down syndrome. *Journal of Autism and Developmental Disorders, 24*, 603-617.

Stanton-Chapman, T. L., Justice, L. M., Skibbe, L. E., & Grant, S. L. (2007). Social and behavioral characteristics of preschoolers with specific language impairment. *Topics in Early Childhood Special Education, 27*(2), 98-109.

Statistical data and indicators on Saudi human resources for 2006-2009. Retrieved May, 2013 from http://www.cdsi.gov.sa/

Stern, D. N. (1985). The interpersonal world of the human infant. New York: Basic Books.

Stiles, J. (1994). On the nature of informant judgments in inventory measures:… and so what is it you want to know? *Monographs of the Society for Research in Child Development, 59*(5), 180-185.

Stone, W. L., & Yoder, P. J. (2001). Predicting spoken language level in children with autism spectrum disorders. *Autism, 5*(4), 341-361.

Stone, W. L., Lemanek, K. L., Fishel, P. T., Fernandez, M. C., & Altemeier, W. A. (1990). Play and imitation skills in the diagnosis of autism in young children. *Pediatrics, 86*(2), 267-272.

Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: What's the object? *Journal of Abnormal Child Psychology, 25*(6), 475-485.

Straus, S., Glasziou, P., Richardson, M., & Haynes, B. (2011*). Evidence-based medicine; how to practice and teach it.* London: Churchill Eivingstone Elsevier.

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences, 10*(10), 471-476.

Suen, H. K., Logan, C. R., Neisworth, J. T., & Bagnato, S. (1995). Parent-professional congruence: Is it necessary? *Journal of Early Intervention, 19,* 243–252.

Swinkels, S. H. N., Dietz, C., van Daalen, E., Kerkhof, I. H. G. M., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum in children aged 14 to 15 months. I: The development of the early screening of autistic traits questionnaire (ESAT). *Journal of Autism and Developmental Disorders, 36*(6), 723-732.

Taylor, L. J., Maybery, M. T., Grayndler, L., & Whitehouse, A. J. (2014). Evidence for shared deficits in identifying emotions from faces and from voices in autism spectrum disorders and specific language impairment. *International Journal of Language & Communication Disorders,* (In press).

Thal, D. J., & Tobias, S. (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. *Journal of Speech, Language and Hearing Research, 35*(6), 1281-1289.

Thal, D., & Bates, E. (1988). Language and gesture in late talkers. *Journal of Speech, Language, and Hearing Research, 31*(1), 115-123.

Thal, D., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: A 1-year follow-up. *Journal of Speech, Language and Hearing Research, 34*(3), 604-612.

Thorp, D., & Mundy, P. (2010). Observational and parent report measures of social communication skills in young children with autism. Unpublished Work

Thurm, A., Lord, C., Lee, L., & Newschaffer, C. (2007). Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(9), 1721-1734.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. J. Dunham. *Joint attention: Its origins and role in development* (pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. C. Levinson (Eds.). *Language Acquisition and Conceptual Development* (pp. 132-158). Cambridge, UK: Cambridge University Press.

Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology, 30*(5), 639-650.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454-1463.

Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology, 17,* 563-584.

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language, 23*(01), 157-176.

Toth, K., Munson, J., Meltzoff, A. N., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders, 36*(8), 993-1005.

Trevarthen, C. & Hubley, P. (1978). Secondary intersubjectivity: confidence, confiding and act of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 183-229). London: Academic Press.

Uzgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development, 4*(1), 1-12.

Van de Vijver, Fons JR, & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 39-63). Mahwah, NJ: Lawrence Erlbaum Associates.

Van der Paelt, S., Warreyn, P., & Roeyers, H. (2014). Social-communicative abilities and language in preschoolers with autism spectrum disorders: Associations differ depending on language age. *Research in Autism Spectrum Disorders, 8*(5), 518-528.

Vanvuchelen, M.,Roeyers, H.,& DeWeerdt,W.(2011a). Development and initial validation of the Preschool Imitation and Praxis Scale (PIPS). *Research in Autism Spectrum Disorders,5*(1), 463–473.

Vivanti, G., Nadig, A., Ozonoff, S., & Rogers, S. J. (2008). What do children with autism attend to during imitation tasks? *Journal of Experimental Child Psychology, 101*(3), 186-205.

Vukovic, M., Vukovic, I., & Stojanovik, V. (2010). Investigation of language and motor skills in Serbian speaking children with specific language impairment and in typically developing children. *Research in Developmental Disabilities, 31*(6), 1633-1644.

Walden, T., & Ogan, T. (1988). The development of social referencing. *Child Development, 59,* 1230–1240.

Wallan, A., Chiat, S., & Roy, P. (2011, November). *Evaluation of an Arabic sentence-repetition test for preschool children*. Poster presented at the annual convention of the American Speech, Language and Hearing Association San Diego, CA.

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language, 108*(3), 175-183.

Watt, N., Wetherby, A., & Shumway, S. (2006). Prelinguistic predictors of language outcome at 3 years of age. *Journal of Speech, Language, and Hearing Research, 49*(6), 1224-1237.

Westerlund, M., Berglund, E., & Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? evaluation of a screening version of the MacArthur-bates communicative development inventories. *Journal of Speech, Language and Hearing Research, 49*(2), 237-247.

Wetherby, A. M., & Prizant, B. M. (1993). *Communication and Symbolic Behavior Scales: Normed Edition.* Baltimore, MD: Paul H Brookes Publishing.

Wetherby, A. M., & Prizant, B. M. (2002). *Communication and Symbolic Behavior Scales: Developmental Profile.* Baltimore, MD: Paul H Brookes Publishing.

Wetherby, A. M., & Prizant, B. M. (2003). *CSBS manual: Communication and Symbolic Behavior Scales.* Baltimore, MD: Paul H Brookes Publishing.

Wetherby, A. M., Allen, L., Cleary, J., Kublin, K., & Goldstein, H. (2002). Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *Journal of Speech, Language and Hearing Research, 45*(6), 1202-1218.

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders, 34*(5), 473-493.

Wiig, E. M., Secord, W., & Semel, E. (1992). CELF-P, *Clinical Evaluation of Language Fundamentals-Preschool.* New York, NY: The Psychological Corporation.

Wiig, E., Secord, W., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals Preschool-Australian and New Zealand Standardised Edition* (CELF P-2 Australian and New Zealand).

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders, 34*(3), 285-299.

Young, G. S., Rogers, S. J., Hutman, T., Rozga, A., Sigman, M., & Ozonoff, S. (2011). Imitation from 12 to 24 months in autism and typical development: A longitudinal rasch analysis. *Developmental Psychology, 47*(6), 1565-1578.

Zanobini, M., Viterbori, P., & Saraceno, F. (2012). Phonology and language development in Italian children: An analysis of production and accuracy. *Journal of Speech, Language, and Hearing Research, 55*(1), 16-31.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale* (4th ed.). San Antonio: The Psychological Corporation.