Strategic verbal rehearsal in adolescents with mild intellectual disabilities: A multi-centre European study

Sebastian Poloczek [a,*], Lucy A. Henry [b,*], Henrik Danielsson [c], Gerhard Büttner [a], Claudia Mähler [d], David J. Messer [e], Kirsten Schuchardt [d], Mariët J. van der Molen [f]

a Department of Psychology, Goethe University Frankfurt and Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany

b Language and Communication Science, City University London, London, UK.

c Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden and The Swedish Institute for Disability Research

d Department of Educational Psychology, Hildesheim University, Hildesheim, Germany

e Faculty of Education and Language Studies, Open University, Milton Keynes, UK

f Department of Clinical, Neuro- & Developmental Psychology, and EMGO Institute for Health and Care Research, VU University, Amsterdam, the Netherlands

* Corresponding authors

E-mail addresses: poloczek@paed.psych.uni-frankfurt.de (S. Poloczek), lucy.henry.1@city.ac.uk (L.A. Henry), henrik.danielsson@liu.se (H. Danielsson), buettner@paed.psych.uni-frankfurt.de (G. Büttner), maehler@uni-hildesheim.de (C. Mähler), david.messer@open.ac.uk (D. Messer), schuchar@uni-hildesheim.de (K. Schuchardt), m.j.vander.molen@vu.nl (M.J. van der Molen)

Keywords: Intellectual disability, memory span, memory strategies, children/adolescents, word length effects, self-pacing

Abstract

Background: There is a long-held view that verbal short-term memory problems of individuals with intellectual disabilities (ID) might be due to a deficit in verbal rehearsal. However, the evidence is inconclusive and word length effects as indicator of rehearsal have been criticised.

Aim & Method: The aim of this multi-site European study was to investigate verbal rehearsal in adolescents with mild ID (*n*=90) and a comparison group of typically developing children matched individually for mental age (MA, *n*=90). The investigation involved: (1) a word length experiment with non-verbal recall using pointing and (2) 'self-paced' inspection times to infer whether verbal strategies were utilised when memorising a set of pictorial items.

Results: The word length effect on recall did not interact with group, suggesting that adolescents with ID *and* MA comparisons used similar verbal strategies, possibly phonological recoding of picture names. The inspection time data suggested that high span individuals in both groups used verbal labelling or single item rehearsal on more demanding lists, as long named items had longer inspection times.

Conclusions: The findings suggest that verbal strategy use is not specifically impaired in adolescents with mild ID and is mental age appropriate, supporting a developmental perspective.

Highlights

- picture memory span of adolescents with mild ID is in line with mental age
- verbal strategy use in adolescents with mild ID is mental age appropriate
- high span individuals use labelling or single item rehearsal on demanding trials
- multi-site investigation increases confidence in the generalizability of findings

What this paper adds?

This paper contributes to the published literature a thoroughly designed and carefully conducted study into the verbal rehearsal of adolescents with ID and chronologically younger, mental age matched children.

To deal with the methodological criticisms of previous studies, we designed a word length experiment that avoided verbal input and verbal responses, but allowed verbal encoding and was in its other methodological features as similar as possible to a verbal serial recall task. However, the key methodological innovation of the paper was that we combined the self-paced presentation method developed by Belmont and Butterfield (1971), but later neglected, with a word length experiment. This combination seems suitable not only to provide further insight into verbal strategy use of adolescents with ID but also to re-examine verbal strategy development in typically developing children.

Additionally, the multi-site collaboration enabled us to collect a large sample for an experimental study involving participants with ID and matched controls. The size and the composition of the sample increased our confidence in the generalizability of the findings and in interpreting non-significant effects with (very) small effect sizes as absence of these effects. Furthermore, the sample size allowed us to gain further insights by examining strategy differences between subgroups with low vs. high memory spans.

1 Introduction

Children with intellectual disabilities (ID) show cognitive and adaptive impairments, and their presence is required for diagnosis (DSM–5; American Psychiatric Association, 2013). Those with ID represent 2-3% of the population (Volkmar & Dykens, 2002), and ID has been described as 'the most handicapping of the disorders beginning in childhood' (Harris, 2006). Nevertheless, the amount of research on children with non-specific ID lags behind that on other childhood disorders (Hodapp & Dykens, 2009) and is below that expected based on prevalence and severity (Bishop, 2010). The current multi-centre European study involved a large sample of adolescents with non-specific aetiology mild ID (IQ between 55-75 and special needs provision) to add to the ecological validity and generalisability of the research. Our investigation addressed a long-standing, pivotal debate about young people with ID, their use of voluntary memory strategies in short-term memory tasks.

Individuals with ID have a wide range of memory difficulties (Henry, 2012; Vicari, 2011). A consistent finding is their impaired immediate verbal serial recall (Henry & MacLean, 2002; Henry & Winfield, 2010; Hulme & Mackenzie, 1992; Russell, Jarrold, & Henry, 1996; Schuchardt, Gebhardt, & Mähler, 2010; Van der Molen, Van Luit, Jongmans, & Van der Molen, 2007). This is found with chronological age-matched peers, but, more significantly, with mental age (MA) matched peers (e.g., Henry & MacLean, 2002; Henry & Winfield, 2010; Hulme & Mackenzie, 1992; Russell et al., 1996). This suggests that impairments in verbal short-term memory (STM) are beyond that expected from MA, and therefore could be a prominent cognitive impairment in individuals with ID (Henry, 2012). Indeed, verbal STM does not appear to develop significantly after 9 years in children with mild and borderline ID (Van der Molen, Henry, & Van Luit, 2014), and poor verbal STM could have important implications for academic achievement, particularly reading and spelling (see Henry & Winfield, 2010; Poloczek, Büttner, & Hasselhorn, 2012), and targeted interventions (e.g., Van der Molen, Van Luit, Van der Molen, Klugkist, & Jongmans, 2010).

Explanations for the verbal STM deficit in these children have focused on the memory trace (weaker in amplitude and duration, e.g., Ellis, 1963), and the characteristics of the 'control' or strategic processes used to improve recall (e.g., Bray & Turner, 1986). Many have argued that children with ID fail to use voluntary memory strategies, and this hypothesis has been supported by experimental findings concerning verbal rehearsal. For example, Belmont and Butterfield (1971) used an innovative procedure whereby lists of six letters were presented under a self-directed pace, controlled by the participant: The authors reasoned that participants using verbal rehearsal would choose to pause longer between items at later serial positions, to allow time for the list items to be recited cumulatively. In fact, adolescents with ID did not pause longer as they went further into the lists, whereas typical adolescents did, suggesting the absence of verbal rehearsal in those with ID.

Hulme and Mackenzie (1992), similarly, argued that poor verbal STM and an absence of word length effects in those with ID (MAs of 5 to 6 years) reflected the absence of strategic verbal rehearsal, or, rephrased in the context of the working memory model (Baddeley, 1986), a deficiency in the 'articulatory rehearsal mechanism'. In this model, verbal information is stored in a time-limited 'phonological store' with only approximately 2-seconds of capacity. However, information can be maintained for longer if the contents are refreshed using subvocal rehearsal via the articulatory rehearsal mechanism. Supporting evidence comes from word length effects when long-named items are recalled more poorly than short-named items in immediate serial recall (e.g., Baddeley, Lewis, & Vallar, 1984). This is assumed to reflect the longer time required for rehearsal of long versus short names.

Other studies have reported an absence of word length effects in children with mild/borderline ID and a MA of around 7 years, compared to mental and chronological age-matched typical children (Hasselhorn & Mähler, 2007; Rosenquist, Conners, & Roskos Ewoldsen, 2003). However, there could be methodological issues with verbal recall methods. Cowan et al. (1992) and Henry (1991) have argued that the process of verbal recall, particularly if it follows auditory presentation of a list, *produces* word length effects, because the recall requires a recitation of the output list and there is more decay in the time-limited phonological store of long named items (see Henry, 2012).

Russell et al. (1996) addressed these issues by asking children to point to pictures of items on a response board, instead of verbally recalling items as in previous experiments. Children with ID and typically developing children (TD) matched for mental age both showed word length effects, but these were small in the non-verbal output condition. The authors were reluctant to interpret their findings as reflecting verbal rehearsal use amongst either group of children. Instead, they suggested that their participants, with MAs of 5 to 7 years, were employing a less sophisticated verbal naming strategy, a view supported by the absence of correlations between articulation rates and memory span in both groups. Jarrold, Baddeley, and Hewes (2000), using similar logic, utilised a probed recall task which eliminated verbal output, and contrasted this with full serial output in children with MAs of 3 to 6 years. They found word length effects only for the serial verbal output condition, again suggesting that verbal output was associated with word length effects. Further, there were few differences between those with and without ID in the apparent use of verbal rehearsal. Finally, Poloczek, Büttner, and Hasselhorn (2014) found a small word length effect in children with ID, which they attributed to output effects. There was a larger effect in a mental age-matched group (6-8 years); this was present at the first serial position and increased with serial position. The authors attributed this finding to an additive effect of output plus verbal rehearsal. Overall, these more recent findings support a ‘deficit’ model of verbal rehearsal in those with ID.

However, such a conclusion may be premature, as not all studies have found evidence of a rehearsal deficit (Schuchardt, Maehler, & Hasselhorn, 2011; Van der Molen et al., 2007) and it remains possible that some individuals with ID may develop the use of verbal rehearsal strategies. The current study examined word length effects as potential indicators of verbal rehearsal in a large sample of adolescents with mild ID drawn from several European countries, with the aim of enhancing the reliability and generalisability of the findings. Careful controls for potential verbal output effects were included, and analyses looked at performance by span level, to determine if higher performing individuals might show evidence for rehearsal, even if lower performing individuals did not. Items for immediate serial recall were presented as pictures with readily available names, rather than as verbal items. We assumed that only children adopting a voluntary strategy of converting pictures into verbal labels ('phonological recoding' – a necessary first step before verbal rehearsal can be employed) would utilise these names as they tried to remember the lists. One set of items to be recalled had long names and one set had short names.

Further, a non-verbal recall method, pointing to items on a pictured array, was employed. This has been used successfully with children with ID in previous research (Henry, 2008; Henry & Winfield, 2010) and provides firmer evidence of phonological recoding and verbal rehearsal because explicit verbalisation is absent, so evidence of phonological recoding and verbalisation should reflect participants' voluntary strategies (see Henry, Messer, Luger-Klein, & Crane, 2012).

Importantly, the mental age of participants in the current study was beyond the proposed start of verbal rehearsal development in typical children (i.e. 6-9 years, Henry, 2012). According to the 'developmental model' (Zigler & Balla, 1982), verbal rehearsal would be expected to develop at the appropriate mental age in adolescents with ID, namely around 6-9 years. A 'difference' perspective on the other hand, would argue for an absence of verbal rehearsal because underlying structural differences in cognitive processes lead to different ways of carrying out memory tasks (e.g., Ellis & Cavalier, 1982).

A further condition was included to provide converging evidence about verbal rehearsal. The Belmont and Butterfield (1971) self-paced presentation was compared to the more usual fixed presentation of lists; picture inspection times of the to-be-remembered items were used to obtain an additional, behavioural strategy indicator of verbal rehearsal. We investigated whether there were increased pause times in self-paced presentation, particularly towards the end of lists when cumulative verbal rehearsal processes would be expected to be more time consuming. We also examined whether pause times were longer for pictures with long as opposed to short names, as would be expected based on the relationship between articulation time and verbal STM.

In summary, the current study investigated the use of verbal rehearsal in adolescents with mild ID, comparing them to chronologically younger MA-matched typically developing children. The key research question was: Do adolescents with mild ID use verbal rehearsal strategies in the same manner and at the same mental age as TD children? A secondary question was: Does the evidence for verbal rehearsal differ in higher span versus lower span individuals?

We predicted that the use of verbal rehearsal in adolescents with mild ID would lag behind that of MA peers, based on literature showing verbal STM impairments and (often) verbal rehearsal deficits in children with ID. However, we tentatively predicted that if evidence for verbal rehearsal in participants with ID was found, this would be for the higher span individuals. We also predicted that typical children would be more likely than adolescents with mild ID to show longer pause times on self-paced lists (to allow more time for verbal recitation of list items), and longer pause times on trials with long words, although such findings might be more apparent for higher span individuals.

2 Method

The study was carried out in three countries (Germany, Netherlands, UK) using exactly the same methodology. The sample size of 90 individuals in each group (ID, MA), to our knowledge, exceeds the largest ever reported on this topic, and the inclusion of participants from three European countries in one multi-centre study further enhances the validity and generalisability of the findings.

2.1 Participants

Students were recruited who had the relevant first language of their country. For the mild ID group, students were recruited from special schools or special educational units within mainstream schools. The ID group was composed of 11- to 18-year-old participants (mean age 14 years 8 months). IQ was assessed using a measure of non-verbal IQ with instruction manuals in all three languages, the Standard Progressive Matrices (SPM, Raven & Horn, 2009; mean nonverbal IQ 65.3; range 55 to 75). Only participants with no identified specific syndrome were included, so that the ID group had mild ID of a non-specific aetiology.

For the MA group, six to nine-year-olds without reported developmental problems were recruited from mainstream elementary schools. Children were included if their SPM raw score matched that of a participant with ID from the same country, and if their nonverbal IQ was between 85 and 115. To calculate IQ the same most recent norms were used (see footnote to Table 1 for details). Since there were no SPM norms for six-year-olds, they were included if they obtained a SPM raw score of at least 17 which is the cut-off for an IQ of 85 in seven-year-olds (this represented a conservative approach). The two groups consisted of 2 x 90 participants, with 33 pairs from Germany, 36 pairs from the Netherlands, and 21 pairs from the UK.

Ethical approval was obtained from each relevant University. Written informed consent from parents/guardians and verbal assent from participants were inclusion criteria.

2.2 Procedure and Materials

Participants were tested at their schools in three sessions. During the first session of approximately 30 min, Raven's SPM was administered in groups of up to five students. The second and third sessions were administered individually on a computer. In the second session (approximately 10 min), a baseline measure of picture span was determined. The third session consisted of the memory experiment (described below) and lasted approximately 25 min.

2.2.1 Word sets. Three sets of nine items were constructed with items chosen to minimise language differences. Each of the 27 items was a familiar, concrete, highly imageable object with a low age of word acquisition. Additionally, 18 objects had short, monosyllabic names in English, Dutch, and German, while 9 objects had long, three- or four-syllable names in all three languages. The words for the baseline set were *arm/arm/Arm, ball/bal/Ball, bed/bed/Bett, book/boek/Buch, cow/koe/Kuh, dog/hond/Hund, fish/vis/Fisch, tent/tent/Zelt,* and *tree/boom/Baum* (*M* age of acquisition (AoA) on a scale of 1 to 7 = 1.69, range 1.25 to 3.05; *M* imageability on a scale of 1 to 7 = 6.48, range = 6.05 to 6.80). The short named objects for the memory experiment were *bread/brood/Brot, chair/stoel/Stuhl, clock/klok/Uhr, clown/clown/Clown, comb/kam/Kamm, glass/glas/Glas, ring/ring/Ring, skirt/rok/Rock,* and *train/trein/Zug or Bahn* (M_{AoA} = 2.13 [1.80 to 2.55]; $M_{\mathrm{imageability}}$ = 6.18 [5.95 to 6.45]). The set of long named words for the experiment consisted of *bikini/bikini/Bikini, elephant/olifant/Elefant, gorilla/gorilla/Gorilla, helicopter/helicopter/Hubschrauber, pineapple/ananas/Ananas, potato/aardappel/Kartoffel, telephone/telefoon/Telefon, umbrella/paraplu/Regenschirm,* and *wheelbarrow/kruiwagen/Schubkarre* (M_{AoA} = 2.55 [2.00-3.20]; $M_{\mathrm{imageability}}$ = 6.26 [5.85-6.70]). Ratings were from the English norms by Morrison, Chappell, and Ellis (1997). Only for *bikini* were no ratings available, but the item was included given the constraints on appropriate items/words in all three languages.

Figure 1. Examples of pictures from the three word sets.

2.2.2 Baseline picture span. This was assessed using the first 'baseline' set of pictures with short names. During instruction, participants were shown all objects on a laptop screen. The experimenter named all objects once to ensure that participants heard the intended labels for items and were more likely to use these if they wished to name the objects. However, the experimenter did not encourage participants to name the pictures or to repeat the names to avoid inducing verbal memory strategies.

The memory task was explained with two pictures. The first picture appeared for 1.5s in the centre of the screen. After an interstimulus interval of 0.5s with a blank screen, the second picture appeared for 1.5s in the screen centre. Thus, pictures did not appear in different spatial positions. Half a second after the last picture was presented, the response array was displayed, with all nine items from the stimulus set in a 3 × 3 array. Participants were asked to indicate the pictures they had seen by pointing to them in the same serial order. As the response array included all items from a particular item pool, item and order memory was required to correctly recall the picture sequences. Verbal responses were not accepted to avoid encouraging verbal coding, and to avoid output timing differences between short and long words. Five different response arrays with the nine objects in random order were used and these were alternated randomly to prevent participants from learning the spatial locations of the items. All of these methodological features (timing of presentations and recall, avoiding spatial cues, serial order report) made the task used during baseline testing and during the experiment as similar as possible to a verbal serial recall task, with the key difference that verbal input and output are avoided.

Assessment of baseline picture span began with lists of one item, and longer lists were presented thereafter until the participant had reached his/her span level. Up to six trials were presented at each list length, but if four trials were entirely correct, trials 5 and 6 were omitted and scored as correct. Whenever participants remembered at least four out of six trials of a particular list length correctly, testing continued with lists lengthened by one item and up to six further trials. Testing was discontinued if three or more trials at a list length were incorrect. Span was defined as the longest list length for which at least four out of six trials were repeated completely correctly. Span-plus-one was defined as one above span level. ‘Low’ span individuals were defined as those achieving spans of 2 or less; and ‘high’ span individuals were defined as those achieving spans of 3 or more.

2.2.3 Strategy self-reports. Before running the word-length experiment, participants were introduced to five possible strategies in picture span tasks. These were: no strategy; a visual strategy; naming (naming each picture once); cumulative rehearsal (repeating previously presented pictures and adding the current picture name); and complete rehearsal (repeating all picture names at the end). After each trial participants were asked to indicate on a response screen with picture cues which strategy they just used. Analyses of self-reports are not reported in this paper.

2.2.4 Word length experiment with fixed presentation rate. In this condition the number of words per trial depended on the participant’s baseline performance. This was to ensure the tasks supported strategy use by not being too easy or difficult, since strategy use may break down when trials are too demanding (Lehmann & Hasselhorn, 2007). Individual titration involved using baseline performance with half the trials at span level and half the trials at span-plus-one level.

As in the baseline task, the experimenter presented and named all pictures in the picture sets with short and long names, and practice trials were administered with each picture being presented for 1.5s and with 0.5s interstimulus intervals. Then, the short-named and long-named objects trials were alternated. The experiment started with six trials at span level and continued with six trials at span-plus-one level. It was recorded which items were remembered in the correct position, so that the proportion of items recalled correctly could be calculated as the dependent variable.

2.2.5 Word length experiment with self-paced presentation rate. In this condition participants decided when to view the next picture. They were instructed to place their finger on the space bar and to press it when they wanted to view the next picture. The time between the start of the presentation of each picture and the next keystroke was logged in milliseconds and gave inspection times for each picture. Times of 300ms or less were set to missing as they were faster than those reported in recognition reaction time experiments. Such very short reaction times often resulted from pressing the spacebar too long. For each individual, very long inspection times of more than 15s and all inspection times below or above 2.5 standard deviations from the individual's mean inspection time were also set to missing, as these probably included atypical response processes.

The self-paced and the fixed presentation conditions were identical in all other respects, including the scoring method. For all participants, the fixed condition preceded the self-paced condition. The rationale for this was to prime a moderately fast-paced presentation, since the aim was to examine strategic behaviour in 'typical' immediate serial recall tasks.

2.2.6 Reliabilities of the dependant variables. Dependent variables included proportion correct recall performance in the fixed and self-paced conditions, and for short and long names. In the self-paced condition only, we recorded inspection times for the short and long sets at two difficulty levels (span, span-plus-one) on lists of up to three or four serial positions (note that this varied between participants depending upon baseline picture span). Internal consistencies of the proportion correct recall performance scores were as follows: short words fixed Cronbach's $\alpha = .59$, long words fixed $\alpha = .60$, short words self-paced $\alpha = .65$, and long words self-paced $\alpha = .72$. Moderate reliability estimates were expected because the trials were adjusted to individual's baseline picture span and therefore inter-individual differences in span could not contribute to the reliability estimates. The internal consistencies of the 14 mean inspection times for the different conditions and positions ranged from $\alpha = .60$ to $\alpha = .79$ with an average reliability of .69.

3 Results

Initial ANOVAs with country as additional between-subject factor revealed no significant interaction effects with country, therefore all data were pooled. The result section is organised with three subsections concerning: (1) whether the ID and MA groups differed in baseline picture span; (2) whether the groups differed in the word length effect, and related analyses concerning whether there were effects of task difficulty, of 'high' or 'low' span, and of self-paced picture presentation; and (3) whether self-paced inspection times provided evidence of verbal rehearsal strategies.

3.1 Baseline Picture Span

Mean picture spans on baseline testing were 2.69 ($SD = 0.82$) for participants with mild ID and 2.86 ($SD = 0.74$) for MA children, with scores ranging from 1 to 6. The group difference in span was not significant, $t(178) = 1.43$, $p = .15$, $d = 0.22$. There were more participants with ID ($n = 36$) than MA children ($n = 22$) who had a span of 2 and were consequently assigned to the 'low span' category; and there were more MA children ($n = 68$) than participants with ID ($n = 54$) who had a span of 3 and consequently were assigned to the 'high span' category.

3.2 Recall in the Word Length Experiment

Mean proportion correct recall performance and standard deviations for the different experimental conditions and groups are provided in Table 2. A five-factor mixed ANOVA with repeated within-participant factors of word length (short vs. long), difficulty (span vs. span-plus-one) and presentation condition (fixed vs. self-paced), and between-participant factors of group (ID vs. MA) and span level (low vs. high) was carried out.

There was a significant main effect of word length, $F(1, 176) = 48.77$, $p < .001$, partial eta squared (η_p^2) = .22, resulting from better recall of pictures with short names. There was also a significant main effect of difficulty, $F(1, 176) = 330.72$, $p < .001$, $\eta_p^2 = .65$, with lower recall in trials at span-plus-one. There was no effect of presentation condition. The main effect of group was significant, $F(1, 176) = 3.97$, $p = .048$, $\eta_p^2 = .022$, due to a slightly higher proportion of correctly recalled pictures for participants with ID. This, however, does not mean that their memory performance was superior, because task difficulty depended on baseline picture span. Finally, the between subjects main effect of span level was significant, $F(1, 176) = 34.78$, $p < .001$, $\eta_p^2 = .16$, with a lower proportion of correct responses for participants with higher spans in baseline-testing (they received longer memory lists).

Of particular interest were the interactions between word length and other variables, as these are relevant to the use of verbal rehearsal. There was a significant interaction between word length and span level, $F(1, 176) = 4.12$, $p = .04$, $\eta_p^2 = .023$, which was due to a larger word length effect for participants with higher spans. There was also a significant interaction between word length and difficulty, $F(1, 176) = 5.78$, $p = .02$, $\eta_p^2 = .032$, due to a larger word length effect on the more difficult lists. However, the word length x group interaction was small and non-significant, $F(1, 176) = 0.77$, $p = .38$, $\eta_p^2 = .004$.

Of the remaining 21 interaction effects, 20 were non-significant with very small effect sizes of η_p^2 = .000 - .010. Only the presentation condition x difficulty x group x span level interaction was marginally significant, $F(1, 176) = 3.90$, $p = .05$, $\eta_p^2 = .022$, and given the large number of analyses, will not be considered further.

The ANOVA findings can be summarized as follows. A word length effect was found, suggesting that the pictures were phonologically recoded. The absence of a group x word length interaction suggested that the ID and MA groups did not differ in their encoding and maintenance of the to-be-remembered pictures. The word length effect was stronger in span-plus-one trials and stronger for participants with higher spans, suggesting different forms of encoding and maintenance for more difficult trials, and for participants with higher spans. However, both these interaction effects were also consistent with the possibility that word length exerts a proportional cost on recall performance that means it is larger at higher span levels, and therefore differences in the word length effect need not be indicative of differences in subvocal rehearsal (e.g., Jarrold & Citroën, 2013; Jarrold, Danielsson, & Wang, 2015). The analyses of the inspection time patterns provide evidence as to which of the two explanations are more likely.

3.3 Self-paced Inspection Times

The inspection times for each serial position in the self-paced experiment were examined to test whether inspection times increased at later serial positions, suggestive of cumulative rehearsal strategies. Since participants with different baseline picture spans were given differing numbers of pictures per trial (list lengths were titrated to ability) this resulted in differing numbers of inspection time variables. Further, different span levels might result in different strategic behaviour during encoding, so inspection time data were analysed separately for participants with a span of 2 ($n_{ID} = 32$, $n_{MA} = 18$) and those with a span of 3 ($n_{ID} = 45$, $n_{MA} = 57$), and separately for trials at span level and trials at span-plus-one level. No ANOVAs were performed for participants with a span of 1 or of 4 and higher, because sample sizes were too small. This resulted in four three-factor mixed ANOVAs on inspection times with the repeated factors of picture type (short, long) and serial position (1 to up to 4) and the between-participant factor of group (ID, MA).

For participants with a span of 2, mean inspection times for each serial position are provided in Figure 2; for those with a span of 3, they are depicted in Figure 3. All ANOVAs revealed a significant main effect of serial position: span of 2 at span level, $F(1, 48) = 65.39$, $p < .001$, $\eta_p^2 = .58$; span of 2 at span-plus-one level, $F(1.38, 66.04) = 56.89$ [1], $p < .001$, $\eta_p^2 = .54$; span of 3 at span level, $F(1.34, 132.79) = 77.16$ [1], $p < .001$, $\eta_p^2 = .44$; and span of 3 at span-plus-one level, $F(2.21, 218.49) = 49.53$ [1], $p < .001$, $\eta_p^2 = .33$. These serial position effects resulted from longer inspection times for the first picture; there was no trend for longer inspection times in later serial positions. The main effects of picture type were small and non-significant for participants with a span of 2 at span level, $F(1, 48) = 0.06$, $p = .80$, $\eta_p^2 = .001$, and at span-plus-one level, $F(1, 48) = 1.68$, $p = .20$, $\eta_p^2 = .034$; as well as for participants with a span of 3 at span level, $F(1, 99) = 0.98$, $p = .32$, $\eta_p^2 = .01$. Additionally, in these three ANOVAs, all other effects including the interaction effects of picture type and group were non-significant and very small ($\eta_p^2 = .000 - .036$), indicating that inspection times for pictures with short vs. long names were similar in both groups.

Only the ANOVA on the inspection times for span-plus-one trials in the participants with a span of 3 revealed a different picture. There was a significant main effect of picture type, $F(1, 99) = 11.20$, $p = .001$, $\eta_p^2 = .10$, due to longer inspection times for pictures with long names. This effect did not interact with group ($\eta_p^2 = .01$). No other effects were significant. One potential criticism of this analysis is that the effect of picture type was significant because of more data (data about a fourth serial position were included). Therefore, an additional ANOVA was run on data from positions 1 to 3. Results were similar, and the main effect of picture type remained significant, $F(1, 99) = 9.97$, $p = .002$, $\eta_p^2 = .09$.

[1] Huynh-Feldt correction applied due to violation of sphericity

Even reducing the sample size to 50 randomly selected participants to give the same power as with the span of 2 data, did not eliminate the significant effect of picture type, $F(1, 48) = 5.11$, $p = .03$, $\eta_p^2 = .10$. Therefore, the effect of longer inspection times for long words was robust for participants with a span of 3 at the more demanding span-plus-one level.

4 Discussion

Our key research question was whether or not adolescents with mild ID used verbal rehearsal strategies in the same manner and at the same mental age as younger MA-matched TD children (6 to 9 years of age). Therefore, the mental age range was higher than in some previous studies on this topic, to ensure that we targeted the age range during which verbal rehearsal develops in typical children. We found a substantial word length effect, which did not interact with group, suggesting that participants with ID and MA comparisons used phonological recoding to convert the pictures into 'verbal picture names', and perhaps also verbally rehearsed the names. Note that given the controversies surrounding interpretation of word length effects, it is not possible from these results to conclude that verbal rehearsal took place (e.g., Henry et al., 2012; Jarrold & Hall, 2013). However, the findings suggest that ID and MA comparisons did not differ in their verbal encoding and maintenance strategies of the to-be-remembered pictures. The large samples from three different European countries should increase confidence about the validity and generalisability of the findings. Further, a carefully designed experimental task with pictures as stimuli and a non-verbal response procedure was utilised to avoid the risk of finding spurious evidence for verbal strategies.

The fact that evidence for phonological recoding and possibly some form of verbal rehearsal was found in both groups at the mental age level expected (e.g., Henry et al., 2012) is contrary to arguments that individuals with ID fail to use any form of verbal memory strategies (Bray & Turner, 1986; Hulme & Mackenzie, 1992). Even if we take a conservative view that only phonological recoding of visually presented items took place, and not necessarily verbal rehearsal, such a strategy still required individuals to voluntary name pictured items, covertly or overtly. The current findings, therefore, provide new evidence regarding strategic behaviour and word length effects in adolescents with mild ID, supporting a ‘developmental’ perspective (Zigler & Balla, 1982) that at least some aspects of verbal strategy development proceed in line with mental age. Importantly, the current findings were not subject to the potentially confounding effects of verbal output at recall or auditory presentation (e.g., Poloczek et al., 2014; Russell et al., 1996).

A subsidiary research question in the current study concerned whether the evidence for verbal rehearsal differed in higher versus lower span individuals. There was some evidence to support this view: participants with higher span levels showed larger word length effects than participants with lower span levels. This interaction effect suggests that rehearsal was used to a greater extent in higher span individuals in both groups. However, Jarrold and Citroën (2013) have cautioned that phonological recoding effects (phonological similarity, word length effects) are subject to proportional scaling. On this argument, the higher the memory span, the more likely we are to find the effects. Some evidence from our data supported their position: word length effects were stronger on trials that were presented at span-plus-one, consistent with the claim that word length exerts a proportional cost on recall performance. Alternatively, participants may have used different forms of encoding and maintenance for more difficult memory trials, and/or participants with higher spans could have adopted different encoding and maintenance strategies.

To provide further evidence about verbal rehearsal, we included a self-paced presentation condition following Belmont and Butterfield (1971). Longer inspection times for pictures with long names versus short names would suggest that phonological recoding is occurring, and longer inspection times towards the ends of lists would imply the use of verbal cumulative rehearsal (more time for verbal recitation of the ever lengthening list). For individuals with a span level of two and three pictures, there was no evidence that inspection times increased with serial position or with word length, and there were no group differences in performance. However, in the more difficult span-plus-one condition, high span individuals showed an effect of word length, with longer inspection times for pictures with long names. This effect did not interact with group, so the findings can be interpreted as evidence for the use of phonological recoding and some form of verbal rehearsal in both groups of high span individuals (ID and MA).

Although this finding suggests the use of a verbal strategy in higher span individuals on more demanding trials, we nevertheless failed to find increasing pause times throughout the list, which would have been an indicator of cumulative verbal rehearsal (Belmont & Butterfield, 1971). Therefore, participants in both groups were probably not using cumulative rehearsal, which requires the entire list to be recited after each new item. Instead, participants were more likely using a simple labelling strategy whereby each picture was named (covertly or overtly) as it was presented, with longer named pictures requiring longer time for inspection/labelling. Such verbal labelling strategies can be regarded as simple verbal strategies that fall short of full cumulative rehearsal, but nevertheless involve a voluntary verbal strategy use (Henry, 2012). The question of whether adolescents or adults with ID who have mental ages above 9 years eventually develop and use cumulative rehearsal in the same manner as TD children do (e.g., Lehmann & Hasselhorn, 2007) cannot be answered by the current findings. However, as verbal STM does seem to reach an early developmental plateau at around 10 years of age in children and adolescents with mild to borderline ID (Van der Molen et al., 2014), one might be sceptical that cumulative rehearsal develops fully in individuals with mild ID.

Furthermore, the development of rehearsal is likely to be complex, variable and nuanced. For example, Lehmann and Hasselhorn (2007) noted many types of verbal behaviour when children remember lists of words/pictures, including labelling (saying the item's name once), single-word rehearsal (repeated one item's name several times), and cumulative rehearsal (repeating at least two words together). These strategies not only vary with age, but individual children use different strategies for different items within the same memory list. Similarly, McGilly and Siegler (1989) found that children rehearsed shorter lists repeatedly, once or not at all, and that behaviour could vary on different trials. Therefore, our tentative conclusion is that higher span individuals in the current study mainly adopted some form of verbal labelling or single-word rehearsal strategy (our method cannot distinguish between these forms of strategy use) on more demanding lists that were beyond their comfortable span. This suggests that the higher span of these individuals could partly be attributable to strategy use and that especially lower span individuals could benefit from a strategy training (e.g., Broadley, MacDonald, & Buckley, 1994).

One final issue concerns the fact that baseline picture span did not significantly differ in the ID and MA groups. Fewer adolescents with ID were classified as having 'higher' spans, but overall span scores did not differ significantly, and, given the large sample size, this is likely to be a reliable finding or any effect is a small one. Picture memory span is similar to verbal STM except that the latter provides a verbal memory trace without recoding. Previous findings suggest a deficit in individuals with ID compared to MA comparisons (e.g., Henry & MacLean, 2002; Henry & Winfield, 2010; Hulme & Mackenzie, 1992; Russell et al., 1996; but also see Conners, Carr, & Willis, 1998; Hasselhorn & Mähler, 2007; Jarrold & Baddeley, 1997), consequently group differences in the baseline span might have been expected. We failed to detect a significant group difference; this could be due to similar verbal STM abilities in both groups or because the two memory tasks involve different abilities (storage versus storage + recoding).

5 Conclusions

The current findings suggest that adolescents with mild ID and MA comparisons with typical development aged 6 to 9 years do not differ in terms of verbal strategy use on a picture memory span task. There was evidence that participants used phonological recoding (word length effects in recall performance), but these findings were the same for both groups. Further, the inspection time data suggested that high span individuals in both groups used some form of verbal labelling or single item rehearsal strategy on more demanding lists. Overall, these findings do not support previous suggestions that verbal strategy use is impaired and fundamentally different in individuals with ID (e.g., Bray & Turner, 1986; Hulme & Mackenzie, 1992), rather that their development of verbal strategy use is merely delayed and in line with their mental age. However, the present study is restricted in its conclusions to individuals with mild ID with a somewhat restricted mental age range. Whether the conclusion of delayed verbal strategy development extends to people with moderate or severe ID, or to children, adolescents or adults with lower or higher mental ages cannot yet be established. Nevertheless, combining a word length experiment with self-paced presentation provides a suitable method to address this question in future research.

Acknowledgements

We would like to thank all children and adolescents for participating in this study, the students at the different universities for assisting in data collection, as well as Bert van Beek for programming all computerized tasks. [***Re-insert Name of Author3 for publication***]'s work in this study was financed by a grant awarded to him from The Swedish Research Council for Health, Working Life and Welfare, FAS 2010-0739.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Baddeley, A. D. (1986). *Working memory*. Oxford: OUP.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36A*, 233–252.

Belmont, J. M., & Butterfield, C. E. (1971). Learning strategies as determinants of mental deficiencies. *Cognitive Psychology, 2*, 411–420.

Bishop, D. V. M. (2010). Which neurodevelopmental disorders get researched and why? *PLoS ONE 5(11)*: e15112.

Bray, N. W., & Turner, L. A. (1986). The rehearsal deficit hypothesis. In N. R. Ellis & N. W. Bray (Eds.) *International review of research in mental retardation*, (Volume 14, pp. 47–71). Orlando, Fl.: Academic Press.

Broadley, I., MacDonald, J., & Buckley, S. (1994). Are children with Down syndrome able to maintain skills learned from a short-term memory training program? *Down Syndrome Research and Practice, 2*, 116–122.

Conners, F. A., Carr, M. D., & Willis, S. (1998). Is the phonological loop responsible for intelligence-related differences in forward digit span? *American Journal on Mental Retardation, 103*, 1–11.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*, 1–17.

Ellis, N. R. (1963). The stimulus trace ad behavioural inadequacy. In N. R. Ellis (Ed.), *Handbook of mental deficiency: Psychological theory and research* (pp. 134–158). New York: McGraw-Hill.

Ellis, N. R., & Cavalier, A. R. (1982). Research perspectives in mental retardation. In E. Zigler & D. Balla (Eds.) *Mental retardation: The developmental difference controversy* (pp. 121–152). Hillsdale, NJ: Lawrence Erlbaum Associates.

Harris, J. C. (2006). *Intellectual disability: Understanding its development, causes, classification, evaluation and treatment*. New York: Oxford University Press.

Hasselhorn, M., & Mähler, C. (2007). Phonological working memory of children in two German special schools. *International Journal of Disability, Development and Education*, *54*, 225–244.

Henry, L. A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *Quarterly Journal of Experimental Psychology*, *43A*, 35–52.

Henry, L. A. (2008). Short-term memory coding in children with intellectual disabilities. *American Journal on Mental Retardation*, *113*, 187–200.

Henry, L. A. (2012). *The development of working memory in children*. London: SAGE Publications Ltd.

Henry, L. A., & MacLean, M. (2002). Working memory performance in children with and without intellectual disabilities. *American Journal on Mental Retardation*, *107*, 421–432.

Henry, L. A., Messer, D. J., Luger-Klein, S., & Crane, L. (2012). Phonological, visual and semantic coding strategies and young children’s short-term picture memory span. *Quarterly Journal of Experimental Psychology*, *65*, 2033–2053.

Henry, L. A., & Winfield, J. (2010). Working memory and educational achievement in children with intellectual disabilities. *Journal of Intellectual Disability Research, 54*, 354–365.

Hodapp, R. M., & Dykens, E. M. (2009). Intellectual disabilities and child psychiatry: Looking to the future. *Journal of Child Psychology and Psychiatry, 50*, 99–107.

Hulme, C., & Mackenzie, S. (1992). *Working memory and severe learning difficulties*. Hove: Lawrence Erlbaum Associates.

Jarrold, C., & Baddeley, A. D. (1997). Short-term memory for verbal and visuospatial information in Down Syndrome. *Cognitive Neuropsychiatry, 2*, 101–122.

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (2000). Verbal short-term memory deficits in Down Syndrome: A consequence of problems in rehearsal? *Journal of Child Psychology and Psychiatry, 41*, 233–244.

Jarrold, C., & Citroën, R. (2013). Reevaluating key evidence for the development of rehearsal: Phonological similarity effects in children are subject to proportional scaling effects. *Developmental Psychology, 49*, 837–847.

Jarrold, C., Danielsson, H., & Wang, X. (2015). Absolute and proportional measures of potential markers of rehearsal, and their implications for accounts of its development. *Frontiers in Psychology, 6*(299).

Jarrold, C., & Hall, D. (2013). The development of rehearsal in verbal short-term memory. *Child Development Perspectives, 7*, 182–186.

Lehmann, M., & Hasselhorn, M. (2007). Variable memory strategy use in children adapted into a task learning behaviour: developmental changes and working memory influences in free recall. *Child Development, 78*, 1068–1082.

McGilly, K., & Siegler, R. S. (1989). How children choose among serial recall strategies. *Child Development*, *60,* 172–182.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528–559.

Poloczek, S., Büttner, G., & Hasselhorn, M. (2012). Relationships between working memory and academic skills: Are there differences between children with intellectual disabilities and typically developing children? *Journal of Cognitive Education and Psychology*, *11*, 20–38.

Poloczek, S., Büttner, G., & Hasselhorn, M. (2014). Phonological short-term memory impairment and the word length effect in children with intellectual disabilities. *Research in Developmental Disabilities, 35,* 455–462.

Raven, J. C., & Horn, R. (2009). *Ravens Progressive Matrices and Vocabulary Scales. Deutsche Bearbeitung und Normierung*. [German adaptation and standardization]. Frankfurt am Main: Pearson.

Raven, J. C., Raven, J., & Court, J. H. (1999). *SPM Manual (Deutsche Bearbeitung und Normierung von St. Bulheller und H. Häcker).* Frankfurt am Main: Swets & Zeitlinger B.V.

Rosenquist, C., Conners, F. A., & Roskos Ewoldsen, B. (2003). Phonological and visuo-spatial working memory in individuals with intellectual disability. *American Journal on Mental Retardation, 108,* 403–413.

Russell, J., Jarrold, C., & Henry, L. (1996). Working memory in children with autism and with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 37*, 673–686.

Schuchardt, K., Gebhardt, M., & Mähler, C. (2010). Working memory functions in children with different degrees of intellectual disability. *Journal of Intellectual Disability Research, 54*, 346–353.

Schuchardt, K., Maehler, C., & Hasselhorn, M. (2011). Functional deficits in phonological working memory in children with intellectual disabilities. *Research in Developmental Disabilities, 32*, 1934–1940.

Vicari, S. (2011). Memory and learning in intellectual disabilities. In J. A. Burack, R. M. Hodapp, G. Iarocci, & E. Zigler (Eds.), *The Oxford handbook of intellectual disability and development* (2nd ed., pp. 97–108). New York: Oxford University Press.

Van der Molen, M. J., Van Luit, J. E. H., Jongmans, M. J., & Van der Molen, M. W. (2007). Verbal working memory in children with mild intellectual disabilities. *Journal of Intellectual Disability Research, 51*, 162–169.

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 54*, 433–447.

Van der Molen, M. J., Henry, L. A., & van Luit, J. E. H. (2014). Working memory development in children with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 58*, 637–650.

Volkmar, F. R., & Dykens, E. (2002). Mental retardation. In M. Rutter, & E. Taylor, (2002). *Child and Adolescent Psychiatry,* (4th Edition, pp. 697–710). Blackwell Science: Oxford.

Zigler, E., & Balla, D. (1982). Introduction: The developmental approach to mental retardation. In E. Zigler & D. Balla (Eds.) *Mental retardation: The developmental difference controversy* (pp. 3–8). Hillsdale, NJ: Lawrence Erlbaum Associates.

Captions for Figures

Figure 1.

Examples of pictures used in the short and long picture name conditions

Figure 2.

Self-paced inspection times (with 95% CIs) at all serial positions for pictures with short and long names (separate lines): participants with a span of 2 showing span level performance (upper section) and span-plus-one level performance (lower section)

Figure 3.

Self-paced inspection times (with 95% CIs) for pictures with short and long names (separate lines): participants with a span of 3 showing span level performance (upper section) and span-plus-one level performance (lower section)

Table 1.

Participant characteristics

	ID (*n* = 90; 69 % male)		MA (*n* = 90; 53% male)	
	M	(SD)	*M*	(*SD*)
Chronological age (in months)	176.0	(16.9)	91.0	(8.3)
Nonverbal IQ[a]	65.3	(5.9)	97.7[b]	(6.9)
SPM raw scores	27.0	(4.1)	26.7	(4.5)

Note. ID = participants with mild intellectual disabilities; MA = mental age matched TD children; SPM = Raven's Standard Progressive Matrices.

[a] In the ID samples, IQ scores are based on the most recent German norms of the SPM (Raven & Horn, 2009), in the MA sample, scores are based on German norms of 1999 (Raven, Raven, & Court, 1999), as more recent norms do not include values for seven to nine-year-olds.

[b] Based on 74 children with age 7;0 and above for whom exact IQs could be calculated.

Table 2.

Proportion correct recall performance for pictures with short vs. long names in the fixed vs. self-paced condition for both groups (ID, MA), with participants at high and low span levels presented separately and combined

			ID group				MA group			
			short		long		short		long	
			M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
complete	fixed	span	.86	(.20)	.80	(.20)	.80	(.24)	.75	(.23)
sample		span+1	.64	(.25)	.56	(.25)	.61	(.26)	.48	(.26)
	self-paced	span	.86	(.19)	.83	(.21)	.81	(.25)	.71	(.28)
		span+1	.69	(.24)	.57	(.27)	.60	(.30)	.46	(.28)
low span	fixed	span	.93	(.20)	.87	(.18)	.86	(.19)	.86	(.20)
(<= 2)		span+1	.74	(.23)	.68	(.20)	.70	(.22)	.60	(.21)
$n_{ID} = 36$	self-paced	span	.96	(.10)	.93	(.16)	.90	(.18)	.84	(.28)
$n_{MA} = 22$		span+1	.73	(.24)	.67	(.26)	.73	(.24)	.66	(.29)
high span	fixed	span	.82	(.20)	.76	(.21)	.78	(.25)	.71	(.23)
(>= 3)		span+1	.58	(.24)	.48	(.24)	.58	(.26)	.44	(.26)
$n_{ID} = 54$	self-paced	span	.80	(.21)	.77	(.22)	.78	(.26)	.66	(.27)
$n_{MA} = 68$		span+1	.65	(.23)	.51	(.25)	.56	(.23)	.40	(.25)