Training effects in dynamic assessment: A pilot study of eye movement as indicator of problem solving behaviour before and after training

Marco G.P. Hessels, Katia Vanderlinden & Hildalill Rojas

Contrary to traditional intelligence tests, dynamic measures of learning capacity have shown to provide reliable measures of children’s general intellectual abilities and prove to be good predictors of future learning. In the present study we used a computerised version of the Hessels Analogical Reasoning Test (HART) to evaluate the changes in problem solving behaviour of children with and without learning difficulties as a result of training. Such training proves to be necessary as many children do not understand what is expected from them in such tasks and, as a consequence, do not use analogical reasoning to solve them. This affects the construct validity of the measure. The training focuses on the rules and procedures one needs to apply to be able to solve analogical problems, for example, systematic inspection of the matrix and the response alternatives, comparison of the different elements in the matrix and the inference and application of the relations found. In learning tests, it is generally assumed that children learn to engage in the processes needed for analogical problem solving during the training, and that the intra-individual variability in effective use of these processes at post-test is indicative of children’s learning capacity. In this study we show that the training indeed provokes children to engage in the appropriate problem solving processes by analysing their visual behaviour. The data show that the children show more structured inspection patterns, more ‘intelligent’ comparisons and spend more time on encoding the information in the matrix, which confirms the HART’s construct validity after training.

Keywords: dynamic assessment; learning potential; eye movement; construct validity; learning difficulties.

Assessment of children with learning difficulties, or intellectual disability, with traditional intelligence tests (IQ) is often criticised for its lack of reliability and (predictive) validity (e.g. Budoff, 1967; Hessels-Schlatter, 2002; Wishart & Duffy, 1990). The use of intelligence tests implicitly has the assumption that all children have had the same opportunities to learn and have had the same learning experiences. From this perspective, differences in intelligence test scores reflect differences in children’s capacity to profit from their learning experiences and the test, thus, differentiates between high and low learning ability. However, it is well known that learning experiences differ, for instance, by cultural or socioeconomic background. Also, for persons with intellectual disabilities, educational goals are often limited to the teaching of social and practical skills, and the teaching of academic skills and knowledge is largely ignored (Hessels-Schlatter, 2002). Furthermore, intelligence tests deliver information about products of former learning opportunities, but do neither provide any information about current learning, nor about the cognitive processes involved. Consequently, the results of future learning of students with learning difficulties (or ‘at risk’ populations in general) are not well predicted and the test results are not useful for the construction of cognitive interventions (Hessels-Schlatter, 2002; Resing, 2006).

Dynamic measures of learning capacity, on the contrary, have shown to provide both reliable and valid measures of the general
intellectual abilities of children with learning difficulties or intellectual disabilities and prove to be good predictors of future learning (Hessels & Hessels-Schlatter, 2010; Resing, 1997). These instruments, also called learning tests or tests of learning potential (LP), use different procedures, but they all have in common that learning is included in the test. An LP test, thus, not only measures what a student has learned so far, but also how easily the student can acquire new knowledge. Research (e.g. Beckmann, 2001; Hessels & Hessels-Schlatter, 2002; Resing, 1993) has shown that when both a traditional IQ test and an LP test are used to predict school-related learning outcomes, the latter not only predicts the same variance as the IQ test, but adds a significant portion of explained variance to the prediction. Beckmann’s (2001) study with adolescents showed that IQ and LP explained about 40 per cent of the same variance in a learning outcome and that LP added 16 per cent to the prediction. Hessels and Hessels-Schlatter (2002), in a study regarding the long-term prediction of learning in minority children aged 5 to 8 years, reported 12 per cent of common variance between IQ and LP and an additional 19 per cent of variance explained by LP. A study with 15- to 18-year-old students with mild intellectual disabilities by Tiekstra, Hessels and Minnaert (2009) showed that IQ predicted a mere nine per cent of variance in a learning outcome (statistically not significant), but that LP raised the prediction to 48 per cent. In all three studies, IQ was not significant when LP was introduced into the equation. The aforementioned authors link the common variance of IQ and LP to the actual level of the student, while the additional explained variance is assumed to be related to the learning capacity of the student (see also Hessels & Hessels-Schlatter, 2010).

The Hessels Analogical Reasoning Test (HART; for example, Hessels, 2009; Hessels, Berger & Bosson, 2008; Tiekstra et al., 2009) is a test of learning capacity which comprises a training phase, followed by a test. Its usefulness for classification and prediction has been shown in various populations, such as immigrant children, students with learning difficulties, and students with intellectual disabilities. The analogies in the HART (as in many other tests of analogical reasoning) are presented in a matrix, where the participant has to discover the rule(s) that govern(s) the relationship between the picture on the left and the one on the right in the first line. The relationship has to be applied to the figure on the left in the second line to find the missing figure on the right. The correct solution is presented below the matrix, amongst a series of possible responses. An example of such a matrix is presented in Figure 1.

The training of the HART focuses on the rules and procedures one needs to apply to be able to solve analogical problems, such as systematic inspection of the matrix and the response alternatives, comparison of the different elements in the matrix and the inference and application of the relations found. Such training proves to be necessary as (young) children, but also students with learning difficulties or intellectual disabilities, often do not understand what is expected from them in such tasks (see, for example, Hessels & Hessels-Schlatter, 2008). Moreover, students with learning difficulties and/or intellectual disabilities tend to show a very passive and incomplete exploration of cognitive tasks, as well as superficial and deficient encoding of the information (Ellis & Dulaney, 1991; McConaghy & Kirby, 1987; Paour, 1992; Wong, 1996). Consequently, they do not engage in analogical reasoning to solve them. Wishart and Duffy (1990) have, for instance, clearly evidenced the unreliability of the Bayley Scales of Infant Development (BSID; Bayley, 1969) tests scores of children with Down’s syndrome. Although the overall scores did not vary much from the first to the second session, many of these children failed items during the second test session which they had previously succeeded. Hessels (2009) reported a lack of consistency in young children’s (5- to 7-year-olds) responses, as well as those
of students in special education classes, to a static pre-test of the HART. The children would, for example, solve some very difficult items, but fail very easy items. Such test results are generally indicative of guessing behaviour instead of problem solving behaviour. These findings not only relate to the reliability of the measures, but also affect their construct validity (Beckmann, 2001, 2006; Hessels, 2009; Hessels-Schlatter, 2002).

When applying learning tests, it is assumed that children learn to engage in the processes needed for (analogical) problem solving during the training, and that the inter-individual variability in effective use of these processes at post-test is indicative of children’s learning capacity. However, the only indicators that are generally used to demonstrate a change are the mean differences found between pre- and post-test scores (using a control group to eliminate practice effects) and the relatively low correlation between pre- and post-test (Tzuriel & Klein, 1985; Budoff, 1967). The latter indicates that the participants did not profit in the same way from the training; some may have gained many points (showing a high learning capacity) and some may have shown no or little gain (showing a low learning capacity). As a result, the order of the participants, in terms of the number of correct responses on the test, changes from pre- to post-test: the greater the differences, the lower the correlation.

An interesting way to examine changes in analogical problem solving as a result of training is to link the performance in such tasks with visual search behaviour and focalised attention by means of eye movement analysis.

**Analogical reasoning and eye movement analysis**

The eye movements we continually make to inspect pictures, to read or to look for an object are called saccades (Rayner, 1998). Saccades are not pre-programmed and can be characterised as non-linear dynamic movements of the eyes (Van der Stigchel, Meeter & Theeuwes, 2006). They are also very rapid, as they can obtain velocities as fast as 500 degrees per second and take about 50 to 150 milliseconds (ms). Saccades allow changing fixation points, thus modifying our gaze, so one can look for relevant information in the visual field. In between saccades, the eyes remain relatively still during 200 to 300 ms, depending on the task being executed. These periods, during which the eyes are positioned on the same point, are called fixations. It has been shown that the sequential distribution of fixations is indicative of the cognitive processes applied to solve a certain cognitive task (Dillon, 1985). According to Yarbus (1967), the order and fixation times of the elements that compose an object are determined by the reflection that accompanies the analysis of the incoming information. Finally, the link between attention and eye location in complex information processing tasks is probably a strong one (Rayner, 1998).

In the scientific literature we find many eye movement studies related to reading and visual perception, but relatively few studies are known in which eye movement analysis was used in relation to analogical reasoning (for a comprehensive overview of eye movement research, see Rayner, 1998).

Bethell-Fox, Lohman and Snow (1984) used eye movement analysis to show that participants (high school students) use different problem solving strategies, depending on item difficulty and the reasoning capacity of the participant, with lower ability participants showing more response elimination (by comparing the elements in the matrix with the alternatives) whereas high ability participants used more elaborate strategies like constructive matching (analysing the information in the analogical matrix to make a mental image of the needed response which is then compared to the response alternatives). According to the authors, item difficulty is related to the number of elements to consider as well as the number of response alternatives, which can lead to memory over-
load. Contrary to Sternberg (1977), they did not find that more able participants used more time to encode the matrix.

A study by Vigneau, Caissie and Bors (2006), applying 14 Raven Advanced Progressive Matrices items (APM; Raven, 1965a) amongst university students, revealed two factors that can explain individual differences in solving analogy problems: speed and strategies. High performing participants were those who spent much time inspecting the analogy matrix (showing relatively long fixation times and saccades within the matrix only), i.e. encoding it completely (see Sternberg, 1977), before their first saccade to the response alternatives. Vigneau et al.’s findings concur with those of Bethel-Fox et al. (1984): the way an individual approaches the matrix and the alternatives, by response elimination or constructive matching, is an important factor for success. Indeed, Vigneau et al. found that participants who spent relatively much time inspecting the matrix, used this information to find the correct response alternative. The authors also found that these participants used less time for encoding items that were perceived as easy, thus speeding up the problem solving process.

Carpenter, Just and Shell (1990) analysed eye movements in participants (university students) who were solving APM items (Raven, 1965a). Carpenter et al. affirm that the strategies for encoding and inducing the regularities in each problem are common to all participants and that individuals are primarily distinguished by their capacity to induce abstract relations and to manage the various steps in problem solving in working memory. It must be noted that the participants in this research were asked to talk out loud while solving problems, which will most probably have influenced their problem solving behaviour (see, for example, Bethge, Carlson & Wiedl, 1982; Carlson & Wiedl, 1979).

Bethge et al. (1982) investigated the impact of dynamic testing on Raven’s Coloured Progressive Matrices (CPM; Raven, 1965b) in third-grade children. They divided their sample into three groups: the first receiving the CPM according to standard instructions; the second received elaborated feedback, i.e. the participants were told whether their answer was correct or not and they were given detailed information why the response was correct or not; the third group was asked to describe the matrix before a response was chosen and to explain why they had chosen a certain response alternative. Bethge et al. found that visual behaviour changed, revealing that procedural characteristics of problem solving changed from one condition to the other. Fixation time and the number of fixations increased in the dynamic versions (i.e. the second and third group), as well as the number of comparisons between the matrix and the distracters, and amongst the distracters. Test performance rose accordingly. The data also suggest that more appropriate and planned strategies were induced by the dynamic procedures.

Dillon (1985) applied Bethge et al.’s procedure with the APM in a group of adult students, to investigate the effects of dynamic procedures on information processing. Dillon found that the processing resources devoted to rule relevant activities increased under the dynamic conditions and she concluded that dynamic procedures have a positive impact on thought regulation during problem solving.

In a recent study, Vakil et al. (2010) investigated eye movements with a computerised version of Tzuriel and Galinka’s (2000) conceptual and perceptual analogical modifiability test (CCPAM) among typically developing (TD) children and adults with mild to moderate intellectual disability (ID)1. The groups were matched on cognitive level (Raven Standard Progressive Matrices (SPM;

---

1 Moderate intellectual disability is similar to severe and complex learning difficulties in the UK.
The authors concluded that significant differences existed between the groups in number of correct scores and time used for the CCPAM, as well as qualitative differences in problem solving, in favour of the TD group. For instance, the proportional number of switches between response alternatives and the proportional number of switches between the matrix and the response alternatives was higher in the ID group. These differences were interpreted as resulting from different strategy use, i.e. the typically developing children using constructive matching and the intellectual disabilities group using response elimination. However, the ID group also showed a much greater proportional number of switches within the matrix, especially for the items in which two or three attributes had to be considered. The comparisons made within the matrix are generally considered as encoding the information needed to construct a mental image of the response, which is indicative of constructive matching. An alternative explanation for the differences found between the groups may be that participants with ID are generally impaired in their working memory capacity (as recognised by the authors in their introduction), and thus need to regularly reiterate the comparisons they make to update their memory and keep track of the attributes that they are treating. This means that they are not so much using a different strategy during problem solving, but need to repeat certain steps because of their working memory limitations. Finally, the significant differences found in CCPAM total scores are remarkable, since the groups were matched on analogical reasoning capacity with the SPM (we assume that no training or feedback was provided in this CCPAM-version, but this is not explicitly stated in the text). Consequently, one would expect equal performance in a similar test of analogical reasoning. The authors explain the difference they found by the fact that the instruction at the beginning of the computerised test may have incited the TD group to be less impulsive during problem solving. However, this explanation is in contradiction with the results, as ID participants use much more time to solve the items than the TD participants and show much more comparisons within the matrix, between the response alternatives and between the matrix and the response alternatives. Such behaviour cannot be considered impulsive. In our opinion, there may be a very simple explanation for the difference found. Firstly, it must be noted that the mean performance of 16 items correct on the SPM in the TD group (with a mean age of exactly 7 years), corresponds to a mean percentile below 25 on this test. This means that the TD group shows below average performance and should probably be considered as a group with borderline intellectual disabilities. Secondly, the computerised CCPAM was presented after the SPM, and even though the modality of presentation is different, it can be considered as a simple retest: both tests address analogical reasoning in the same kind of format (pictorial or figural). It is well known that performances rise due to simple test repetition for average or borderline children, but not for persons with moderate to severe ID (Hessels-Schlatter, 2002; Schlatter, 1999; Wishart & Duffy, 1990). It also shows that matching for cognitive level using a static intelligence test is probably not the most appropriate way to achieve equal groups when working with persons with ID (see Hessels & Gassner, 2010).

Aim of the present study
The aim of the present study is to explore whether the training in the HART, aimed at procedural aspects of analogical reasoning, can change problem solving processes in primary and secondary school students. The problem solving processes may become visible through eye movement recording and by analysing visual behaviour during analogical reasoning tasks. Such a process analysis will not only contribute to the construct validity of the HART, but to that of learning tests in general. We report here a first pilot study in which eye movements of partici-
pants, with and without learning difficulties, during analogical problem solving were recorded, both before and after training.

Method
In this section we will describe the sample, the instruments used and the procedure we applied for testing. We remind the reader that this was a first pilot study, mainly aimed at constructing the computerised tests, to develop an appropriate procedure for measuring eye movements and to prepare a more controlled experimental study. However, even though the study has some evident methodological shortcomings, we think that the outcomes of this small pilot study may be interesting to a larger audience.

Sample
The sample first of all consisted of four students (all girls) with learning difficulties attending the Learning Centre (Atelier d’Apprentissage), a pedagogical service at the University of Geneva. The aim of the Learning Centre is to remediate students’ learning problems through (meta-)cognitive intervention. Two of the students (9- and 10-years-old) were in a mainstream school, the other two (13- and 14-years-old) were in a special educational class. All four students participated in the study before starting the intervention at the Learning Centre. Their mean age at the time of the research was 12 years and 1 month.

To enlarge the sample, the second and third author asked children of friends and acquaintances to participate in the study. Seven students, all without learning difficulties, volunteered to participate in the study. The two girls and five boys were aged 8 (grade 3) to 12 years (grade 7) at the time of the research. Their mean age was 10 years and 7 months.

The data of five other students who participated in the research could not be included in the study because of problems with the eye movement registration.

Materials
C-HART pre- and post-test: To be able to measure analogical reasoning and at the same time register eye movements, we first constructed two parallel computerised versions of the HART: C-HART1 and C-HART2 (Rojas & Vanderlinden, 2010). The 10 parallel items of each version were extracted from the original 66-item set of the paper version of the HART. The parallel versions contained the same type of relations, but applied to different attributes. For the purpose of this research, the original items of the HART which are presented in a 3 (row) x 2 (columns) or 3 x 3 format, with either 6 or 8 response alternatives below the matrix, needed to be adapted to arrive at a suitable screen display. This not only concerned an easy on the eye presentation for the students, but also related to the eye movement registration. Indeed, the display had to allow for enough space between the elements on the screen for clear identification of the zones that were scanned by the student, taking into account the error range of the eye tracker device. This means that the original presentation format was standardised to a 2 rows x 2 columns format with only 4 response alternatives presented below the matrix: (1) a correct response alternative; (2) a nearly correct alternative, with one detail changed or missing; (3) an incorrect response alternative, with several wrong elements or missing parts; and (4) a completely wrong alternative, with irrelevant attributes. The item sizes on the screen were slightly larger than those in the printed version. The example that is used to familiarise the students with the test is presented in Figure 1.

The individual training in the HART was provided using the original paper materials (Hessels et al., 2008) and comprised learning to compare attributes and to infer relations with regard to simple figures, and to use these skills in solving six training items.
Eye movement tracking
The C-HART1 and C-HART2 were integrated into a Tobii 1750 eye tracking system. The cameras, hidden in the 17-inch TFT monitor on which the items are displayed with a resolution of 1280 x 1024 pixels, track the infrared markers that are also mounted on the screen. The system allows computing the gaze position with an accuracy of 0.5°. No head restraint was needed as the eye tracking camera automatically adjusts to head movements. Saccades were detected by the Tobii Clearview system.

Procedure
The students were instructed to remain well seated at a normal distance in front of the computer screen. Before starting testing, the students were asked to look at a dot that was displayed sequentially in the four corners and the centre of the screen to calibrate the positions of both eyes.

The C-HART1 was introduced by an introductory item, using the same type of standard instructions as Raven Matrices. The student then has to solve the 10 test items following his or her own rhythm. Testing time generally takes about 10 minutes.

The training was presented in a second session and took about 15 minutes. The training was immediately followed by the C-HART2 (on this occasion without the introductory item).

Measures
C-HART1 and C-HART2 both provided a score representing the total number of items correct. Both pre-test and post-test appeared to contain one ambiguous item. These were removed from the analyses, reducing the maximum score of each test to 9.

The eye tracking video information first of all allowed the calculation of a series of mean latencies (in milliseconds) across the nine items. We will only present the most important ones in this study: the time used for the first encoding of the matrix (Encoding), the total time used for inspection of the matrix (Matrix), the total time used for inspecting the response alternatives (Responses), the time used to look elsewhere on the screen, outside relevant zones.

Figure 1: Familiarisation item of the computerised versions of the HART.
of the item (Elsewhere), and the total time used for solving the item (Total time). The latencies were further used to calculate the proportional times (in relation to the total time) for Encoding, Matrix and Responses.

The videos further allowed quantifying other visual behaviours, such as inspection of details within the matrix during first encoding (Details), the number of useful comparisons of attributes within the matrix during first encoding (M-Comparisons), and the number of comparisons between the matrix and the alternatives (MA-Comparisons).

**Analyses**
The univariate tests will be one-tailed, as significant increases are expected for C-HART, Encoding, Matrix, proportional time for encoding, proportional time on matrix, Details and M-Comparisons, and significant decreases are expected for Responses, Elsewhere, Total time, Proportional time for responses and MA-Comparisons. Cohen’s *d* will be presented to illustrate effect sizes. An effect size *d* of 0.2 is considered small, a *d* of 0.5 is seen as a moderate effect size and a value of 0.8 is considered a large effect (Cohen, 1988).

As the C-HART2 showed a ceiling effect (small standard deviation and negatively skewed), the difference between C-HART1 and C-HART2 was tested with a non-parametric test (Wilcoxon Z). The effect size was calculated using the formula suggested by Field (2005): Z/√N.

**Results**
Table 1 presents the means and standard deviations of all variables at pre- and post-test, as well as the comparisons of the results before and after training with *t*-tests and effect sizes.

The table first of all shows that the mean test score (C-HART) increases significantly from 6.6 at pre-test to 8.2 at post-test. The effect size is large. As mentioned above, the

<table>
<thead>
<tr>
<th>Table 1: Means and standard deviations of all variables at pre- and post-test, <em>t</em>-tests (or Wilcoxon) and effect sizes.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-test</strong></td>
</tr>
<tr>
<td><strong>M</strong></td>
</tr>
<tr>
<td><strong>Test score</strong></td>
</tr>
<tr>
<td>C-HART*</td>
</tr>
<tr>
<td><strong>Latencies</strong></td>
</tr>
<tr>
<td>Encoding</td>
</tr>
<tr>
<td>Matrix</td>
</tr>
<tr>
<td>Responses</td>
</tr>
<tr>
<td>Elsewhere</td>
</tr>
<tr>
<td>Total time</td>
</tr>
<tr>
<td><strong>Proportional latencies</strong></td>
</tr>
<tr>
<td>Proportion encoding</td>
</tr>
<tr>
<td>Proportion matrix</td>
</tr>
<tr>
<td>Proportion responses</td>
</tr>
<tr>
<td><strong>Behaviours</strong></td>
</tr>
<tr>
<td>Details</td>
</tr>
<tr>
<td>M-Comparisons</td>
</tr>
<tr>
<td>MA-Comparisons</td>
</tr>
</tbody>
</table>

*Wilcoxon Z, effect size r=Z/√N **All latencies in milliseconds, except Total time in seconds

Marco G.P. Hessels, Katia Vanderlinden & Hildalill Rojas

Educational & Child Psychology Vol. 28 No. 2
small standard deviation illustrates that a ceiling effect appears at post-test. This is especially true for participants without learning difficulties in the sample, who all obtained either eight or nine points. Table 1 further shows that the mean latencies significantly decrease for Responses and Total time, but significantly increase for Encoding. The mean latency of Elsewhere changes only slightly (but not significantly) in the expected direction, whereas the change in Matrix is negative. The effect sizes are moderate to large. The increased latency for Encoding means that students take more time to encode the information in the matrix, before showing the first saccade to the response alternatives. The shorter overall time for Matrix is most likely related to the fact that generally less time is needed for the items.

The proportional measures show that the percentage of the time spent on Encoding rises from 10.8 per cent at pre-test to 15.7 per cent at post-test. This difference is significant at the 5 per cent level and the effect size is moderate. The proportion of total time used for inspection of the matrix changes only slightly (from 29.9 per cent to 32.1 per cent). The proportion of time used for the response alternatives (30.2 per cent at pre-test and 24.1 per cent at post-test) also showed a tendency (p≤.10) to decrease. The effect size is moderate.

Finally, the table also shows that two important behaviours change: the number of details inspected during first encoding (Details) and the number of useful comparisons of attributes within the matrix during first encoding (M-Comparisons) increase significantly and the effect sizes can be considered large. The number of comparisons between the matrix and the alternatives (MA-Comparisons) show a small but not significant decrease. The effect size is moderate.

To further investigate the changes in problem solving behaviour, Spearman correlations between the proportional latency measures, behavioural measures and test performance were computed, but for the children with learning difficulties only (Table 2). We chose to do so, because of the ceiling effect that was found in the group of participants without learning difficulties at post-test.

At pre-test, C-HART1 shows a strong and negative correlation with Encoding and Details, a low correlation with Matrix, Responses and M-Comparisons, and a strong positive correlation with MA-Comparisons. Since the number of participants is very small, only the correlation with Details is significant at five per cent level. The correlations imply that a high performance in the HART, amongst the children with learning difficulties, is related to a low proportional inspection time of the matrix, little inspection of details and many comparisons between the matrix and the alternatives. This corresponds to the strategy of Response Elimination.

At post-test, we find low positive correlations with Encoding and Responses, a moderate correlation with Matrix, a very
strong correlation with Details and a perfect correlation with M-Comparisons. A tendency for a strong negative correlation is found for MA-Comparisons \((p \leq .10)\). The latter three correlations, in particular, indicate that a high performance is related to a detailed inspection of the attributes in the matrix, to many comparisons within the matrix (needed to infer the relations) and with few comparisons between the matrix and the response alternatives. This corresponds to using the strategy of Constructive Matching.

**Conclusion**

In this paper, we argued that dynamic measures of learning capacity are more appropriate than traditional intelligence tests when assessing persons with learning difficulties or intellectual disabilities. In particular, the extent to which students profit from the training provided within the dynamic tests is seen as indicative of learning capacity. We hypothesised that the training in the HART, aimed at the procedural aspects of analogical reasoning, would change problem solving processes in students. This hypothesis is related to the construct validity of the learning test. To be able to link test performance of participants with and without learning difficulties to problem solving strategies, we recorded visual search behaviour and focalised attention (eye movements) during analogical problem solving, before and after training. An increasing amount of research evidence shows that eye movement analysis can adequately provide information about problem solving processes. Although some eye movement studies with dynamic assessment exist in which the behaviours of groups of students are compared (Bethge et al., 1982; Dillon, 1985), this is (to our knowledge) the first study in which eye movement is used to evidence the changes in problem solving strategies within students.

The present study not only showed that C-HART scores increased from pre-test to post-test, but also that the participants were much more rapid at post-test. The participants became more efficient problem solvers. The results further showed that although the participants were more rapid overall, they allocated significantly more time to encoding of the matrix before turning their attention to the response alternatives. They also paid more attention to the inspection of details and made more comparisons within the matrix to be able to infer the relationships. These aspects are considered indicative for a Constructive Matching strategy (see Vigneau et al., 2006 and Vakil et al., 2010).

At pre-test, on the contrary, the participants allocated much more time to the responses and also showed more comparisons between the matrix and the responses (although not tested for significance in each group separately because of the small numbers, inspection of the data showed that participants, both with and without learning difficulties, showed these changes in their problem-solving behaviour). This behaviour is seen as characteristic for a Response Elimination strategy. Bethel-Fox et al. (1984), Vigneau et al. (2006) and Vakil et al. (2010) associate such a strategy with low ability participants, whereas the Constructive Matching strategy, based on good encoding, is associated with high ability individuals (see also Sternberg, 1985).

Furthermore, the complete reversions of the correlations from pre- to post-test in the small group of participants with learning difficulties further illustrates the general change from using a Response Elimination strategy at pre-test to using a Constructive Matching strategy at post-test. These observations confirm the hypothesised impact of the training and subsequent construct validity of the HART.

In fact, one could argue that the Response Elimination strategy is not only an inefficient strategy to solve analogical reasoning tasks, but also that it does not involve analogical reasoning at all, since it is not aimed at inferring the rules that govern the transformations between the matrix elements A, B and C, and to apply these rules
to element C to find the missing element D. If we take the argument further, this means that we are measuring analogical reasoning in high ability participants only, since they do apply a Constructive Matching approach to the task. The low ability participants appear to be engaged in qualitatively different processes. Thus, if we want to measure analogical reasoning in all participants, we must assure that they know how to engage in the required processes. This implies that participants should be provided with an appropriate introduction (training) to the task, as is done in dynamic measures of reasoning capacity. In light of the arguments mentioned in the introduction and the results presented in this paper, traditional standard tests without such an introduction may be deemed unsuited for participants with learning difficulties or intellectual disabilities (see also Hessels, 2009).

Of course, this was a small pilot study and we cannot make strong generalisations. The children without learning difficulties who participated could not be differentiated because of ceiling effects at post-test. In further studies, we should, therefore, extend the C-HART with a number of more difficult items. The children with learning difficulties on the other hand could be clearly discriminated and did allow us to make some remarkable observations, as was shown for instance by the shift in correlations. We hope to be able to confirm these findings in an upcoming study with a much greater number of participants.

The relatively simple hardware allows an easy use in a multitude of settings. The eye tracking system with a camera and infrared sensors mounted in the computer screen can be connected to a portable computer and can easily be taken into the classroom or institution. Heavy, immobile equipment is no longer required, and neither is a head restraint as the system automatically corrects for head movements of the participants (within a frame of about 30 x 30 x 30 cm). Such systems may, therefore, allow educational psychologists to analyse problem solving behaviour of children in various computerised tasks within the classroom. Eye movement registration can then become a key element of educational psychology assessment, helping to identify deficient cognitive processes and strategies in children with learning difficulties. Such knowledge is essential for creating specific interventions, tailored to the needs of the individual child (Hessels-Schlatter, 2010).

Acknowledgement
The authors wish to thank Christine Hessels-Schlatter for her valuable comments on an earlier draft of this text.

Address for correspondence:
Marco Hessels
University of Geneva,
Faculty of Psychology and Educational Sciences,
Boulevard du Pont d’Arve, 40,
1205 Geneva,
Switzerland.
Email: Marco.Hessels@unige.ch
References


Marco G.P. Hessels, Katia Vanderlinden & Hildalill Rojas


