Individual differences in the effect of feedback on children's change in analogical reasoning

Claire E. Stevenson
Leiden University, Psychology Methods & Statistics
Wassenaarseweg 52, Postbus 9555
2300 RB Leiden, The Netherlands
+31 71 527 3789
cstevenson@fsw.leidenuniv.nl

Paul A. L. de Boeck
Ohio State University, Dept. of Psychology
232 Lazenby Hall, 1827 Neal Ave,
Columbus, OH, 43210, USA
+1 614 292 4131
deboeck.2@osu.edu

Wilma C. M. Resing
Leiden University, Developmental Psychology
Wassenaarseweg 52, Postbus 9555
2300 RB Leiden, The Netherlands
+31 71 527 3789
resing@fsw.leidenuniv.nl

Willem J. Heiser
Leiden University, Psychology Methods & Statistics
Wassenaarseweg 52, Postbus 9555
2300 RB Leiden, The Netherlands
+31 71 527 3789
heiser@fsw.leidenuniv.nl

ABSTRACT

Various forms of feedback are used in formative assessment and interactive learning environments. The effects of different types of feedback are often examined at a group level. However, effective feedback may differ in learners with different characteristics or between learners at different stages in the learning process. In this paper explanatory item response theory (IRT) models are used to examine individual differences in feedback effects in children's performance on a computerized pretest-training-posttest assessment of analogical reasoning. The role of working memory and strategy-use as well as interactions between these factors were examined in a sample of 1000 children who received either stepwise elaborated feedback, repeated simple feedback or no feedback during the training sessions. The results show that working memory efficiency significantly predicted initial ability and confirm that elaborate feedback is the most effective form of training in this particular interactive learning environment. Furthermore, children with initially less advanced strategy-use benefitted far more from each type of feedback than the children displaying more advanced strategies and this was unrelated to working memory efficiency. In children with advanced strategy-use working memory appears to moderate the effect of training. Explanatory IRT analyses appear useful in disentangling the effects of learner characteristics on performance and change during formative assessment and could possibly be used in optimizing feedback in computerized training and assessment environments.

Keywords

Figural analogies, measuring change, item response theory, formative assessment

1. INTRODUCTION

Computer-based interactive learning environments have enormous potential in optimizing learning by providing feedback tailored to an individual's instructional needs. However, determining what type of feedback best optimizes the learning of a particular task for a particular individual is a complex endeavor. The effectiveness of different types of feedback is not always clear-cut. Furthermore, individual differences may be present in how effective each of these types of feedback is at different stages in the learning process.

In formative assessment different types of feedback can be used. Shute distinguished a range of feedback-types from simple forms such as verification of correct response to elaborated feedback where errors may be flagged, an opportunity to try again is provided and/or strategic prompts are given on how to proceed with the problem [Shutte 2008]. Kluger and DeNisi [1996] argued that although simple feedback, such as information on correctness of response or provision of the correct answer, has the reputation of improving performance on tasks, its effect is not clear-cut and only improves performance or learning in two-thirds of the studies included in their meta-analysis. Furthermore, more recent research demonstrates that elaborate feedback, such as providing scaffolds or an explanation, is generally more effective than simple outcome feedback [Hattie and Gan, 2011; Narciss and Huth 2006; Shutte 2008]. For example, a meta-analysis of effects of different forms of item-based feedback in computer-based environments reports that elaborated feedback shows higher effect sizes than simple outcome feedback, especially in higher-level learning outcomes, where transfer of previous learning to new situations or tasks is required [van der Kleij et al. 2013].

In the case of formative assessment the aim is to optimize learning at an individual level. In this educational setting the assumption is that there are individual differences both in initial ability as well as the effect of different types of feedback during an individual's learning process. Furthermore, different types of feedback may be more effective during successive stages in the learning process. However, effective feedback may differ for different types of learners or at different stages in the learning process. For example working memory efficiency and strategy-use have been implicated as predictors of performance in (computer-based) learning [Siegler and Svetina, 2002; Stevenson 2012; Tunteler et al. 2008]. In this study these factors were examined in conjunction with feedback-type as possible predictors of learning outcomes in a computerized training and assessment of analogical reasoning.

Initial ability or learning stage especially appears to play an important role in the effect of different forms of feedback on learning [Hattie and Timperley 2007]. For example, in a previous study on children's change in analogical reasoning training utilizing repeated simple feedback was contrasted with graduated prompting techniques, a form of elaborated feedback where increasingly specific strategic hints guide the child to the correct solution [Campione and Brown 1987; Resing and Elliott, 2011]. The researchers found that although graduated prompts led to greater performance gains on the whole, this form of training was most effective for children who performed poorly on the pretest [Stevenson et al. 2013a]. These results could not be explained by ceiling effects or regression to the mean. Furthermore, this result coincided with other cognitive training studies in various domains where interventions were generally more effective in initially lower performing or at-risk populations. Does this mean that providing elaborate versus simple feedback is not necessarily beneficial for more advanced learners?

To further explore the role of initial ability on feedback effects we examined the role of children's initial solution strategies (analogical versus non-analogical, see Figure 1) in the effect of three types of feedback: (1) step-wise elaborated feedback, (2) repeated simple feedback or (3) no feedback. The hypothesis was that children with initially weaker analogical reasoning strategies, characterized by "duplicate" (copying object next to empty box) solutions or "other / creating a zoo" solutions would benefit most from more elaborate forms of feedback whereas children who were already capable of applying analogical reasoning strategies (providing (partially) correct solutions) would not show differential benefit in the different types of feedback training. The role of working memory, which has often been shown to be related to analogy solving skills, but not always able to account for children's change in analogical reasoning [Stevenson et al. 2013b], was also taken into account in these analyses.

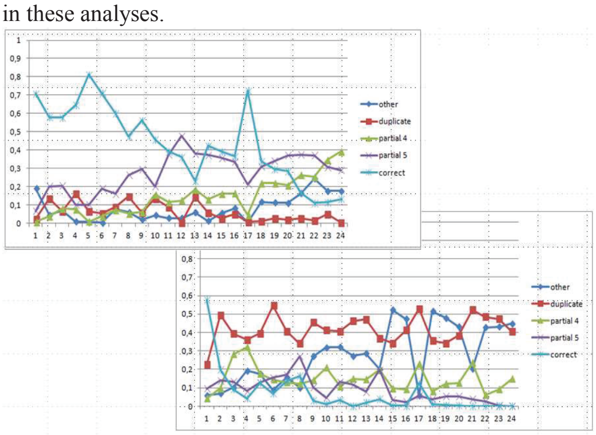

Figure 1. Depiction of strategy distribution within two pretest strategy groups: non-analogical reasoners (top left) and analogical reasoners (bottom right).

2. METHODS

2.1 Sample

1000 children from five age-groups (kindergarten, first through fourth grade) were recruited from public elementary schools of similar middle class SES in the south-west of the Netherlands. The sample consisted of 374 boys and 626 girls, with a mean age of 7 years, 3 months (range 4.9-11.3 years). The schools were selected based upon their willingness to participate and written informed consent for children's participation was obtained from the parents.

2.2 Design & Procedure

The data utilized in this study is a combination from five separate studies utilizing a pretest-intervention-posttest control-group design [Stevenson 2012]. In each study the children were randomly blocked to the step-wise elaborative feedback (graduated prompts), repeated simple feedback or a control condition without feedback based on their scores on a cognitive ability reasoning subtest (visual exclusion from the Revised Amsterdam Children's Intelligence Test [Bleichrodt et al. 1987] or the Standard Progressive Matrices [Raven et al. 2004]). The three intervention conditions presented in this study are: (1) stepwise elaborate feedback, (2) repeated simple feedback, or (3) no feedback. Four analogy testing and intervention sessions took place weekly and lasted 20-30 minutes each. Prior to the analogy testing sessions the children were also administered the Automated Working Memory Assessment to assess verbal (subtest listening recall) and visuo-spatial (spatial span) working memory [Alloway 2007]. All participants were tested individually in a quiet room at the child's school by educational psychology students trained in the procedure.

2.3 Analogical reasoning assessment

AnimaLogica was used to test and train children in analogical reasoning [Stevenson 2012]. The figural analogies (A:B::C:?) comprise of 2x2 matrices with familiar animals as objects (see Figure 2). The animals changed horizontally or vertically by color, orientation, size, position, quantity or animal type. The number of transformations – or object changes – provide an indication of item difficulty [Mulholland et al. 1980]. The children were asked to construct the solution to the analogy using drag & drop functions to place animal figures into the empty box in the lower left or right quadrant of the matrix. A maximum of two animals were present in each analogy. These were available in three colors (red, yellow, blue) and two sizes (large, small). The orientation (facing left or right) could be changed by clicking the animal figure. Quantity was specified by the number of animal figures placed in the empty box. Position was specified by location of the figure placed in the box.

The pretest and posttest items were isomorphs [Freund and Holling 2011] in which the items only differ in color and type of animal, but utilize the exact same transformations to ensure the same difficulty level. The number of items different per age group but included overlapping items ability could be estimated reliably using item response models. The internal consistency of each of the versions was considered very good with $\alpha \geq .90$.

Before each testing or training session two example items were provided with simple instructions on how to solve the analogies. If the child's solution was incorrect the correct solution was shown before proceeding to the next item. During the testing phases the remaining items were administered without feedback.

Table 1.

Overview of the prompts used in the elaborative feedback condition.

Prompt	Verbal Instruction
0	Here’s a puzzle with animal pictures. The animals from this box have been taken away. Can you figure out which ones go in the empty box?
1	Do you remember what to do? Look carefully. Think hard. Now try to solve the puzzle.
2	This animal picture changes to this one. This one should change the same way.
3	So what changes here? Ok remember this one changes the same way.
4	See, this picture changes to this one because…
5	Which animal goes in the empty box? The elephant or the horse? What color should it be? Red, Yellow or Blue? …Size? Quantity? Orientation? Position?...

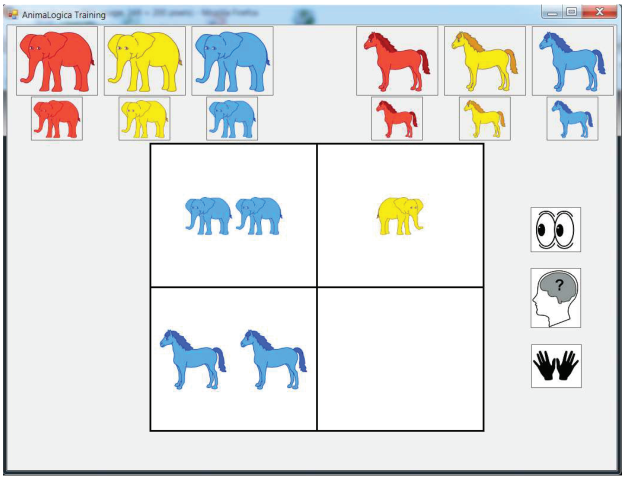

Figure 2. Depiction of visual effects emphasizes cues from prompt 1 to “Look carefully”, “Think hard” and then “Try to solve the puzzle” (these are not all shown at once).

2.3.1. Feedback Interventions.

The *stepwise elaborate feedback condition* received training according to the graduated prompts method [Campione and Brown 1987; Resing and Elliott 2011] which consisted of stepwise instructions beginning with general, metacognitive prompts, such as focusing attention, followed by cognitive hints, emphasizing the transformations and solution procedure, and ending with step-by-step scaffolds to solve the problem (see Table 1). The prompts were mostly auditory in nature and accompanied by visual effects support the explanations (see Figures 2 & 3). A maximum of five prompts were administered. Once the child answered an item correctly the child was asked to explain his/her answer; no further prompts were provided and the next item was administered.

The *simple feedback condition* received auditory feedback on whether or not the outcome was correct and this was repeated until the item was solved correctly or five attempts were made to solve the item. After the fifth incorrect attempt the correct solution was shown before proceeding to the next item. If a correct solution was found before five attempts then the next item was administered.

In the *control condition* the children received the exact same items as in the other two conditions but did not receive help or feedback in solving them. Therefore, the children only practiced solving the items but were not trained in analogical reasoning.

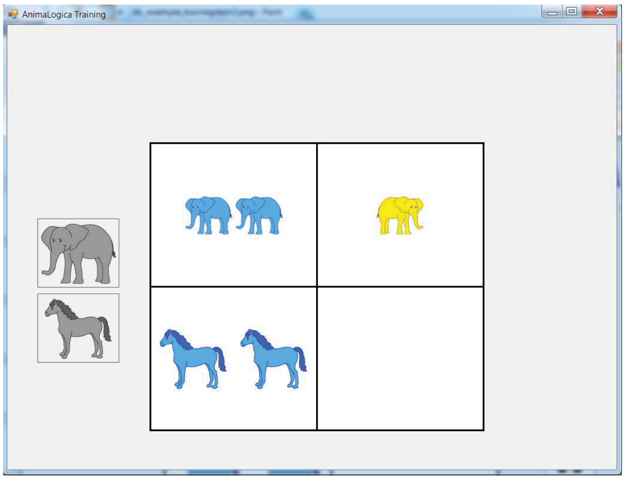

Figure 3a. Visual effects emphasizing prompt 5 where scaffolds are used to solve the puzzle: “Which animal belongs in the empty box?”.

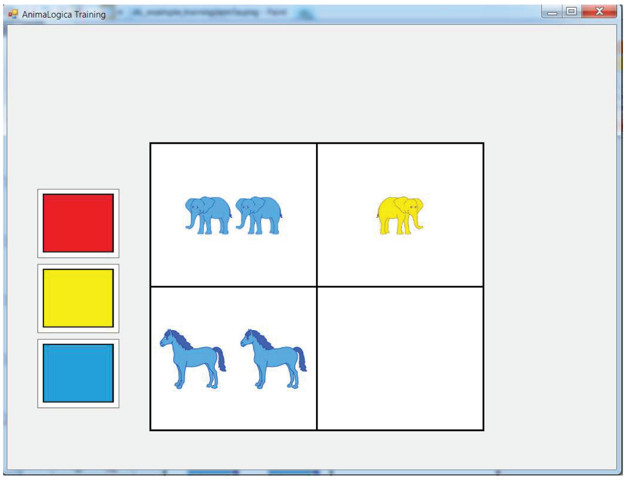

Figure 3b. Prompt 5 scaffold: “What color should it be?”.

2.4 Statistical Models

Disentangling the complex changes in ability over time on an individual basis requires complex statistical models. For example, using raw gain scores (posttest minus pretest score) to measure change can lead measurement errors due to the unreliability of the gain score, the regression effect of repeated administration and that the scale units for change do not share constant meaning for test takers with different pretest scores and [de Bock 1976; Lord 1963]. These problems are potentially solved by placing ability scores for pretest and posttest on a joint interval measurement scale using logistic models such as those employed in item response theory (IRT) [Embretson and Reise 2000]. In the Rasch model, one of the most simple IRT models, the chance that an item is solved correctly depends on the difference between the latent ability of the learner and the difficulty of the presented item or problem. The Rasch-based gain score provides a good basis for the latent scaling of learning and change because the gain score has the same meaning in terms of log odds (i.e. the logarithm of probability of correct vs. incorrect) across the entire measurement scale [Embretson and Reise 2000]. Therefore, this study applied IRT models to analyze individual differences in feedback effects on learning and change [Stevenson et al. 2013a].

2.4.1 Explanatory IRT analyses

Each of the hypotheses about the children's performance and change was investigated using model comparison. First a reference model was created and then predictors were added successively to so that the fit of the new model could be compared to the previous (nested) model using a likelihood ratio (LR) test, which assesses change in goodness of fit. The models were estimated using the lme4 package for R [Bates and Maecheler 2010] as described by [De Boeck et al. 2011].

2.4.2 Null model

The initial reference model (M0) was a simple IRT model with random intercepts for both persons and items (pretest and posttest) where the probability of a correct response of person *p* on item *i* is expressed as shown in equation 1.

$$P(y_{pi} = 1 \mid \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}$$

$$\text{where } \theta_p \sim N(0, \sigma_\theta^2) \text{ and } \beta_i \sim N(0, \sigma_\beta^2) \quad (1)$$

2.4.3 Modelling learning and change

This study employs repeated testing. In order to account for this effect a session parameter has to be added to the null model to represent average change from pretest to posttest. However, this model assumes the effect of retesting to be equal for all children. In order to allow for individual differences in improvement from pretest to posttest a random parameter that allows for the session effect to vary over persons was added. In this model, Embretson's Multidimensional Rasch Model for Learning and Change (MRMLC, see M2 in Table 1), the chance that an item is solved correctly (P_{ip}) also depends on the difference between the examinee's latent ability (θ_p) and the item difficulty (β_i) [Embretson 1991]. Yet, the ability is built up through the testing occasions *m* up to *k* in a summation term, which indicates which abilities (θ_{pm}) must be included for person *p* on occasion *k*.

$$P(y_{ipk} = 1 \mid \theta_{pk}, \beta_i) = \frac{\exp(\sum_m^k \theta_{pm} - \beta_i)}{1 + \exp(\sum_m^k \theta_{pm} - \beta_i)}$$

$$\text{where } \theta_{pm} \sim N(0, \sigma_\theta^2) \text{ and } \beta_i \sim N(0, \sigma_\beta^2) \quad (2)$$

The initial ability factor, θ_{p1}, refers to the first measurement occasion (i.e. pretest) and the so-called modifiabilities (θ_{pm} with $m>1$) represents the change from one occasion to the next. In the present model examining pretest to posttest change $k=2$ and the modifiability θ_{p2} refers to performance change from pretest to posttest.

2.4.4 Modelling sources of individual differences in learning and change

The formula in equation 2 can be extended by including other item or person predictor variables and evaluating their effects on the latent scale [De Boeck and Wilson 2004]. Person predictors are denoted as Z_{pj} *(j=1,…,J)* and have regression parameters ζ_j. The item predictor (e.g. number of transformations) can be denoted as X_i (k=1) and has the regression parameter δ. These predictors are successively entered into the null model (see equation 1) as follows, with indices *i* for items, *p* for persons, *j* for the person covariate used as a predictor variable and *k* for the item covariate used a predictor variable.

$$P(y_{pi} = 1 \mid Z_{p1} \ldots Z_{pJ}, \beta_i) = \frac{\exp(\sum_{j=1}^J \varsigma_j Z_{pj} + \varepsilon_p + \delta X_{ik} + \varepsilon_i)}{1 + \exp(\sum_{j=1}^J \varsigma_j Z_{pj} + \varepsilon_p + \delta X_{ik} + \varepsilon_i)}$$

$$\text{where } \varepsilon_p \sim N(0, \sigma_{\varepsilon p}^2) \text{ and } \varepsilon_i \sim N(0, \sigma_{\varepsilon i}^2) \quad (3)$$

This equation represents models M3-6 in the results presented in Table 2.

Table 2.

Overview of the estimated IRT models.

Model	Nested Model	Effects: Fixed	Effects: Random over Persons	Effects: Random over Items	AIC	BIC	-LL	LR test[a]: *df*	LR test[a]: *Λ*
M0			Intercept	Intercept	37575	37600	18784		
M1	M0	+ Session	"	"	35741	35775	17866	1	1835.90***
M2	M1		+Session	"	34871	34922	17429	2	874.18***
M3	M2	+ Session* Condition	"	"	34063	34132	17024	2	811.52***
M4	M3	* Strategy group	"	"	33773	33944	16866	12	314.50***
M5	M4	* WM	"	"	18014	18236	8979	8	15775***

[a] The LR-test comprises a comparison between the model and the nested model. *** p < .001

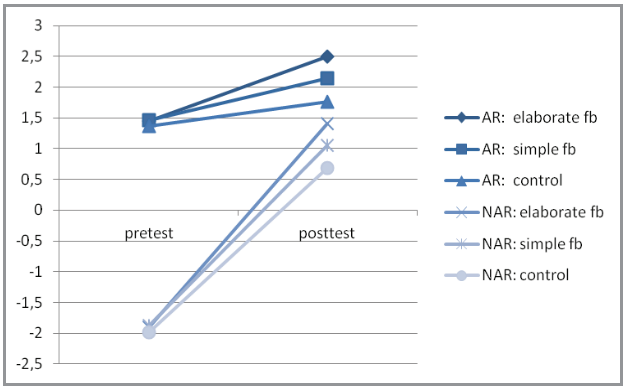

Figure 4. Plot of M5 with logit (y-axis) by Session (x-axis) for Analogical Reasoners (AR) versus Non-analogical reasoners (NAR) for each feedback condition (elaborate, repeated simple and control).

3. RESULTS

Table 2 displays the outcomes of the model building steps. As can be seen in the right-most column the addition of each new predictor in the explanatory IRT model significantly improved model fit. From M0 to M1 we could statistically infer that there was a main effect for training. The inclusion of individual regression lines for performance change from the pretest to posttest was deemed warranted given the improved model fit from M1 to M2. The significant model comparison result from M2 to M3 shows us that the different types of feedback had different "change" slopes. The difference in performance change from pretest to posttest between the two strategy-groups is shown in model M4 (see Figure 4). Finally, from M4 to M5 we could statistically infer working memory was differentially related to performance change per condition and strategy group. Analysis of the simple contrasts indicated that working memory moderated feedback effects in the analogical reasoners (AR strategy group), but was unrelated to performance change in the non-analogical reasoners (NAR strategy group) (simple feedback: $B = -1.38$, $p < .01$ and elaborated feedback: $B = -1.37$, $p < .01$, reference category = no feedback / control condition).

Significant fixed main effects were found for Session, Strategy group, verbal and visuo-spatial Working memory. Significant fixed interaction effects were found for Session x Condition, Session x Strategy group, Session x Working memory, Strategy group x Working memory and Session x Strategy group x Working memory. Random intercepts were present for persons ($SD_{ability} = .62$, $SD_{modifiability} = .70$, $r = -.24$) and items ($SD = .74$).

Table 3.

Estimates of fixed effects in M5.

	B	SE	*p*
Intercept	- 0.32	.42	.44
Session (reference = pretest)	2.17	.16	<.001
Simple Feedback Condition (reference = control)	0.10	.10	.32
Elaborate Feedback Condition (reference = control)	0.08	.10	.41
Strategy-group (reference = non-analogical reasoners)	3.26	.11	<.001
Verbal working memory	0.23	.09	.01
Visuo-spatial working memory	0.26	.04	<.001
Session * Simple Feedback Condition	0.28	.13	.04
Session * Elaborate Feedback Condition	0.65	.13	<.001
Session * Strategy-group	-1.65	.12	<.001
Session * Verbal Working memory	0.47	.11	<.001
Strategy-group * Verbal Working memory	0.08	.10	.43
Session * Strategy-group * Verbal Working memory	-0.61	.13	<.001

4. CONCLUSION

This paper presented our recent research in the area of statistical models of formative feedback effects in performance and change in children's analogical reasoning. The results showed that individual differences stemming from initial strategy-use and working memory efficiency were present and influenced the effect feedback. Elaborate feedback was more effective than simple feedback. Working memory was a predictor of pretest performance. Working memory also moderated feedback effects but only in children in the advanced strategy-use group. Working memory most likely forms a bottleneck in children's analogical reasoning on difficult analogy tasks [Richland et al. 2006]; however children with less advanced strategies most likely were unable to solve the more difficult analogy items which would require accurate solving steps and the accompanying greater taxation of working memory to do so. Finally, initial strategy-use interacted with feedback-type in that children using less advanced strategies at pretest benefited more from each form of feedback during training compared to the children displaying more advanced strategies at pretest. On the whole, the main conclusion is that elaborated feedback, presently implemented using graduated prompting techniques, appears to be the advisable form of feedback in advancing children's change in analogical reasoning.

Given the great potential of computer-based interactive learning environments to provide feedback tailored to an individual's instructional needs an important task is creating algorithms to optimize feedback provision and thus learning. On the one hand (meta-analyses of) randomized pretest-training-posttest control experiments that contrast the effectiveness of different types of feedback and explore sources of individual differences herein as discussed in the present paper provide essential information concerning which factors could be used to optimize feedback. However an investigation of the effects of specific elaborated feedback prompts on a trial-by-trial basis [Golden et al. 2012] and the interactions with learner characteristics or task performance (e.g., strategy-use) using item response theory models is a promising next step towards the provision of optimal feedback in interactive learning environments. Thus the next step in this research project is to expand upon the present findings concerning the effectiveness of the stepwise elaborated feedback and disentangle the immediate effects of the separate prompts during the training process. It will be interesting to see whether different types of prompts better aid more or less advanced learners with more or less efficient working memory to solve the items presented during training.

5. ACKNOWLEDGMENTS

Thank you to Marian Hickendorff for her assistance with the initial analyses of the strategy groups presented in this paper. We also thank Femke Stad and Carlijn Bergwerff for their assistance with data collection and coding.

6. REFERENCES

[1] SHUTE, V. J. 2008. Focus on Formative Feedback, *Review of Educational Research*, 78 (1), 153-189.

[2] KLUGER, A. N. AND DENISI, A.1996. The Effects of Feedback Interventions on Performance: A Historical Review , a Meta-Analysis , and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 2(2), 254-284.

[3] NARCISS, S. AND HUTH, K. 2006. Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, 16(4), 310-322.

[4] HATTIE, J. AND GAN, M. 2011. Instruction Based on Feedback. In *Handbook of Research on Learning*, R. E. MAYER AND P. A. ALEXANDER, eds. New York, New York, USA: Routledge, 249-271.

[5] VAN DER KLEIJ, F. M. FESKENS, R. C. W. AND EGGEN, T. J. H. M. submitted 2013. The effectiveness of methods for providing written feedback through a computer-based assessment for learning: A systematic review. 1-25.

[6] STEVENSON, C.E. 2012. *Puzzling with Potential Dynamic testing of analogical reasoning in children*.

Doctoral dissertation. Amsterdam: Leiden University, 1-191.

[7] SIEGLER, R. S. AND SVETINA, M. 2002. A microgenetic/cross-sectional study of matrix completion: comparing short-term and long-term change. *Child development*, 73(3) 793-809.

[8] TUNTELER, E., PRONK, C. M. E. AND RESING, W. C. M. 2008. Inter- and intra-individual variability in the process of change in the use of analogical strategies to solve geometric tasks in children: A microgenetic analysis. *Learning and Individual Differences*, 18(1), 44-60.

[9] HATTIE, J. AND TIMPERLEY, H. 2007. The Power of Feedback, *Review of Educational Research*, 77 (1), 81-112.

[10] CAMPIONE, J. C. AND BROWN, A. L. Linking dynamic assessment with school achievement. In *Dynamic assessment: an interactional approach to evaluating learning potential*, C. S. LIDZ, Ed. New York, New York, USA: Guilford Press, 82-109.

[11] RESING, W. C. M. AND ELLIOTT, J. G. 2011. Dynamic testing with tangible electronics: measuring children's change in strategy use with a series completion task. *The British journal of educational psychology*, 81(4), 579-605.

[12] STEVENSON, C. E., HICKENDORFF, M. RESING, W. C. M., HEISER, W. J. AND DE BOECK, P. A. L. 2013a. Intelligence Explanatory item response modeling of children's change on a dynamic test of analogical reasoning, *Intelligence*, 41(3), 157-168.

[13] STEVENSON, C. E., HEISER, W. J. and RESING, W. C. M. 2013b. Working memory as a moderator of training and transfer of analogical reasoning in children, *Contemporary Educational Psychology*, 38(3) 159-169.

[14] MULHOLLAND, T. M., PELLEGRINO, J. W. AND GLASER, R. 1980. Components of geometric analogy solution. *Cognitive psychology*, 12(2) 252-284.

[15] FREUND, P. A. AND HOLLING, H. 2011. How to get really smart: Modeling retest and training effects in ability testing using computer-generated figural matrix items, *Intelligence*, 39(4), 233-243.

[16] F. DE BOCK, 1976. Basic issues in the measurement of change. In *Advances in psychological and educational measurement.*, D. N. M. DE GRUIJTER AND VAN DER L. J. T. KAMP, Eds. New York, New York, USA: Wiley.

[17] LORD, F. M. 1963. Elementary models for measuring change. In *Problems in measuring change.*, C. W. HARRIS, Ed. Madison, Wisconsin, USA: University of Wisconsin Press, 21-38.

[18] EMBRETSON, S. E. AND REISE, S. 2000. *Item response theory for psychologists*. Mahwah, NJ: Erlbaum Publishers.

[19] BATES, D. AND MAECHELER, M. 2004, lme4: Linear Mixed-Effects models using S4 Classes. http://r-forge.r-project.org/projects/lme4/.

[20] DE BOECK, P. A. L., BAKKER, M., ZWITSER, R., NIVARD, M., HOFMAN, A. TUERLINCKX, F. AND PARTCHEV, I. 2011. The estimation of item reponse models with the lmer Function from the lme4 Package in R, *Journal of Statistical Software*, 39 (12), 1-27.

[21] EMBRETSON, S. E. 1991. A multidimensional latent trait model for measuring learning and change, *Psychometrika*, 56(3), 495-515.

[22] DE BOECK, P. A. L. AND WILSON, M. 2004. *Explanatory item response models: A generalized linear and nonlinear approach*. New York, New York, USA: Springer.

[23] RICHLAND, L. E. MORRISON, R. G. AND HOLYOAK, K. J. 2006. Children's development of analogical reasoning: insights from scene analogy problems. *Journal of experimental child psychology*, 94(3), 249-73.

[24] GOLDIN, I. M. KOEDINGER, K. R. AND ALEVEN, V. A. W. M. M. 2012. Learner Differences in Hint Processing, In *Proceedings of the 5th International Conference on Educational Data Mining,* Chiana, Greece, 2012.

[25] BLEICHRODT, N., DRENTH, P.J.D., ZAAL, J.N. & RESING, W.C.M. 1987. *Handleiding bij de Revisie Amsterdamse Kinder Intelligentie Test [Manual of the Revised Amsterdam Child Intelligence Test]*. Lisse: Swets & Zeitlinger.

[26] RAVEN, J. RAVEN, J.C. & COURT, J.H. 2004. Manual for Raven's Progressive Matrices and Vocabulary Scales. San Antonio, Texas: Harcourt Assessment.

[27] ALLOWAY, T.P. 2007. *Automated Working Memory Assessment (AWMA)*. London: Harcourt Assessment.