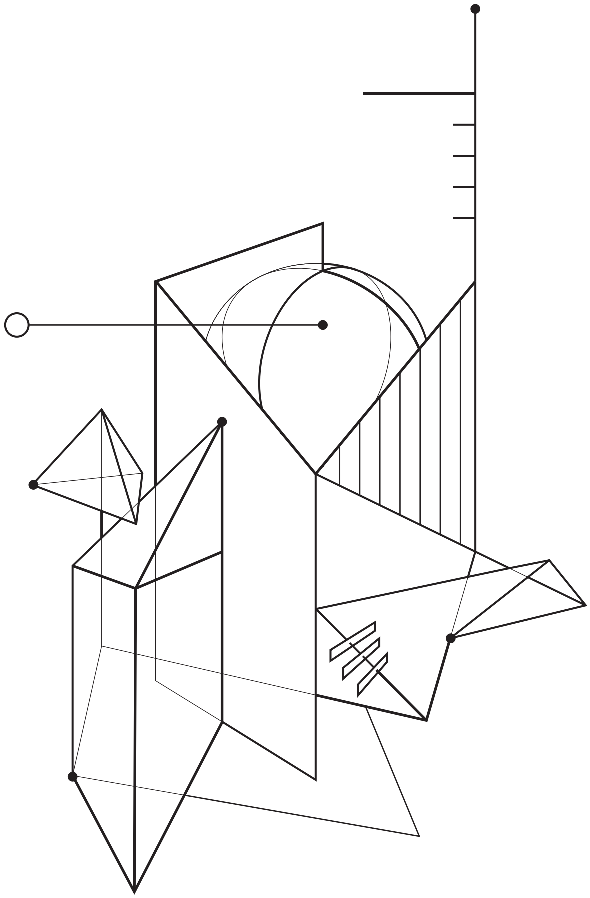
Hidden treasures: Uncovering task solving processes in dynamic testing

Veerbeek, J.

Author: Veerbeek, J.
Title: Hidden treasures: Uncovering task solving processes in dynamic testing
Issue Date: 2019-04-11

CHAPTER 6

General discussion



The main goal of this thesis was to investigate the information that can be obtained from process-oriented dynamic testing using a rule-based, theory-driven scoring method. More specifically, the specific characteristics of Grouping of Answer Pieces (GAP), a method of computer automated process assessment based on children's sequences of solving activities when answering cognitive problems, were object of study.

In this chapter, a summary of the findings from the studies described in this thesis was provided, and the theoretical considerations and practical implications of these findings were discussed. Then, the limitations in our studies were reviewed, and finally recommendations for future research were provided.

Summary of findings

The main aim of the first study presented in this thesis (*Chapter 2*) was evaluating a new measure for analyzing task solving processes in series completion tasks in 8-year-old children. The measure, *Grouping of Answer Pieces* (GAP), was thought to provide information on children's problem representation and restructuring. Information regarding children's task solving processes was collected, utilizing an electronic tangible user interface was. The GAP measure was found to be a moderate predictor of accuracy in solving the series completion tasks, as were verbalized strategy use and time measures. Analyses at the level of individual items revealed that, depending on item difficulty, the GAP measure was the only process measure that was moderately related to successfully solving the task. Furthermore, the GAP measure was not significantly related to verbalized strategy use and time measures, indicating that it measured unique variance. In addition, more stable grouping behavior was related to better series completion solving ability. In contrast, for verbalized strategy use, more variability was related to better task outcomes. Taken together, these findings support the notion that these measures provided information on different aspects of the task solving process.

Chapter 3 focused on the effects of using or not using a pretest in a dynamic series completion task in 7-to-8 year old children. Half of the children received a pretest, a graduated prompts training and a posttest, the other half only received the graduated prompts training and the posttest. No differences were found between children that had received a pretest and children that had not, in terms of series completion solving ability on the posttest, the processes used in solving the posttest items in terms of the GAP measure, verbalized strategy use, and planning time, or the number of hints they needed during training. Further analyses showed that process measures, more specifically the GAP measure and verbalized strategy use, predicted item success. The lack of difference between children that had and had not received a pretest, indicated that the decision of whether or not to use a pretest could be based on the diagnostic questions to be answered and the testing situation, rather than on psychometric considerations.

Chapter 4 investigated 7-to-8 year old children's task solving processes and changes in these processes as a result of a graduated prompts training in a dynamic series completion task. Half of the children received a graduated prompts training, the other half received repeated practice only. Children's task solving processes were measured in terms of the GAP measure, verbalized strategy use, and completion time. Trained children showed more progress in their series completion solving ability from pre- to posttest than untrained children did. In relation to task solving processes, the effects of training were most visible in children's verbalized strategy use, which became more advanced as a result of training. With regard to the GAP measure, trained children did not show a more advanced level of use than untrained children, but they did show different distributions over the different GAP categories. There were no differences in completion time of children who were trained or only received repeated practice. Although the process measures were found to be related to performance on the dynamic series completion task, as well as to math and reading comprehension performance, the amount of help children received during training had most predictive value for school performance.

In *Chapter 5* the use of process-oriented dynamic testing was investigated within a different domain, again using a pretest/training/posttest design, where half of the children received a training, and the other half received repeated practice only. In this chapter, a visual-spatial complex figure task was used to investigate the effects of training on the processes on and outcomes of this task with 7-to-8 year old children. A graduated prompts training procedure was developed, aimed at improving children's organization of the figure. Training led to more progression from pre- to posttest in terms of children's complex figure drawing performance, as well as use of more advanced grouping behavior. These findings indicated a shift in processes used to complete the complex figure drawing task, which seemed to have been brought about by the graduated prompts training. Only trained children were able to attain the highest level of organization of the figure.

Theoretical and practical considerations

The studies of this thesis were focused on process-oriented dynamic testing (e.g., Resing, Bakker, Pronk, & Elliott, 2017; Resing, Xenidou-Dervou, Steijn, & Elliott, 2012; Resing & Elliott, 2011). In-depth dynamic assessment was thought to unveil more information regarding children's instructional needs. In this thesis, process-oriented measurement was found to provide valuable information beyond outcome scores, including the effects of training on a child's task solving processes.

Dynamic testing

The majority of the studies in this thesis (*Chapter 3, 4, and 5*) made use of dynamic testing, which was executed using a graduated prompts training procedure (e.g., Resing, 2013). Dynamic testing was found to lead to significant increases in children's solving performance in both a series completion task (*Chapter 4*), and a complex figure drawing task (*Chapter 5*). These findings were in line with previous research that successfully used the graduated prompts approach in dynamic testing of, among others, series completion ability (e.g., Resing et al., 2012; Stad, Vogelaar, Veerbeek, & Resing, 2017), analogical reasoning (e.g., Stevenson, Heiser, & Resing, 2013), and language learning (e.g., Camilleri & Botting, 2013; Hasson, Dodd, & Botting, 2012). Taken together, these findings seem to support the notion that the graduated prompts training method can lead to improved task performance, irrespective of the task domain. Furthermore, in line with previous research (Caffrey, Fuchs, & Fuchs, 2008; Stevenson, Bergwerff, Heiser, & Resing, 2014; Swanson & Howard, 2005), dynamic testing measures in series completion were found to provide moderately more explained variance than static testing measures

(*Chapter 4*). Particularly the number of hints children needed during training provided a better prediction of children's academic performance than static measures, in line with previous research (e.g., Stevenson et al., 2014).

Despite the promising features of dynamic testing, a question has been raised concerning the influence of using a pretest/posttest design on the validity and reliability of the test scores (Kim & Willson, 2010; Klauer, 1993; Sijtsma, 1993; Sternberg & Grigorenko, 2002). In *Chapter 3*, the test performances of children receiving a full dynamic test (pretest/training/posttest) and children receiving training and posttest only, were compared to address this issue. No differences were found in children's outcomes on series completion tasks, nor in the processes children employed to solve the task. This led us to conclude that for the graduated prompts dynamic series completion test, the use of a pretest did not lead to changes to the reliability or validity of the test. As a result, when working with dynamic testing in practice, when deciding whether or not to use a pretest, the question to be answered by that test might be used in the first place rather than psychometric differences in the construct measured by the test.

Process assessment

Several process measures were utilized to investigate children's task solving processes during dynamic testing in series completion. Children's verbal explanations of their task solving processes were moderately related to accuracy on the series completion task (*Chapter 2, 3, and 4*), and appeared to be most in line with the characteristics of strategy use as described by Siegler (1996, 2007), who defined strategy use as variable, both between and within tasks. Variability in verbalized strategy use was related to greater accuracy in the series completion task (*Chapter 2*), indicating that, in line with the theory of Hunt (1980), rather than being connected to one particular strategy, successful task solving behavior is connected to the ability to adaptively alternate between strategies if the tasks requires this. Children's verbalized strategy use progressed towards more sophisticated explanations as a result of training (*Chapter 4*). Apparently, the learning situation did not only lead to improved performance, but also more advanced strategy use, which was in line with the interactive relationship between strategy use and learning, as described by Alibali, Phillips, and Fischer (2009).

Compared with the other available process measures, completion time appeared to have the weakest relationship with accuracy on the series completion task (*Chapter 2 and 4*), and was not influenced differently by training than by repeated practice (*Chapter 4*). This may have been a result of the level of difficulty of the tasks, which in the past has been shown to influence the relationship between completion time and accuracy (Dodonova & Dodonov, 2013; Goldhammer et al., 2014). The time taken for planning correlated only moderately with accuracy (*Chapter 2*), whereas accuracy was related more strongly to the verbalization and GAP measures.

Grouping of answer pieces (GAP). The grouping of answer pieces in children's task solutions was thought to be another process measure, being indicative of the way in which children represented a problem, and of strategies that were employed to more efficiently store this information, by grouping related information together into "chunks" (Halford, Wilson, & Phillips, 1998; Pretz, Naples, & Sternberg, 2003). In line with the theory of Sternberg (1985), who stated that problem representation is a metacomponent of problem solving, more stable scores on the GAP measure, as an indicator of children's problem representation, were related to better task outcomes. The GAP measure was consistently found to be a moderate predictor of item success (*Chapter 2, 3, and 4*). In line with the literature, the task representation as measured with the GAP measure was closely related to successful task solving (Hunt, 1980; Pretz et al., 2003; Robertson, 2001).

Restrictions of children's task representation have been related to limitations of their working memory capacity. The use of more effective strategies for storing information has been argued to lead to a more efficient mental representation of the task (e.g., Andrews & Halford, 2002; Halford et al., 1998). Halford and colleagues (1998) concluded that if children do not have these strategies available, didactic support is a prerequisite to help children more efficiently represent the information necessary for solving the task. In line with this research, children's representations of the complex figure task as operationalized with the GAP measure, improved as a result of a targeted graduated prompts training procedure, but not as a result of repeated practice (*Chapter 5*). The series completion study revealed a more differentiated picture regarding the different process measures utilized. Trained children profited from the dynamic training given to them and showed more progress in their series completion solving ability from pre- to posttest than their control group peers. Under influence of training, their verbalization measures progressed towards a more advanced level. The GAP measures of both trained and untrained children progressed as well. Both dynamic tests used in this thesis were very different from each other, including the training scripts provided to the children. For the moment, we have to conclude that task content, task difficulty, and variation in training procedures influence the role of GAP measures in process-oriented dynamic testing.

The overall picture that emerged from the findings support the notion that each process measure reveals a different part of the actual solving steps children showed when solving complex cognitive tasks. Grouping behavior, as measured by the GAP measure, certainly represents one aspect of the problem representation as described by Sternberg (1985), and can be thought to be a stable metacomponent, which serves to guide the task solving process (*Chapter 2*). Some children, however, might need a didactic intervention to efficiently use the problem representation (e.g., Halford et al., 1998).

Assessment of children's cognitive abilities. Throughout this thesis, a new measure was used, which would have been very difficult to obtain without the use of computers. The GAP measure was developed from an algorithm which detected the grouping of pieces in the overall sequence of placement or drawing of the answer. The basic principle was developed first for a series completion task, based on extensive task analysis. Later, the same basic principle was redeveloped for a complex figure task, based on previous research of the processes involved in solving complex figure drawing tasks. The use of constructed response items has been advocated in the past, as they provide more and more detailed information on children's task solving abilities and processes (Stevenson, Hickendorff, Resing, Heiser, & de Boeck, 2013; Yang, Buckendahl, Juszkiewicz, & Bhola, 2002). Through combining them with computerized systems, the nature of scoring the tasks becomes less labor-intensive, and, as was found in this thesis, information on children's task solving processes can be obtained relatively easily.

Tangible User Interfaces. Through the use of tangible user interfaces (TUIs), the benefits of computers were combined with the benefits of tangible materials for children's development, and to elicit authentic behavior from the children (Verhaegh, Resing, Jacobs, & Fontijn, 2009). Using such computerized systems enabled children to work independently and receive feedback and training from the console (Resing et al., 2012; Verhaegh, Fontijn, Aarts, & Resing, 2013). It also provided the possibility of log-file analysis as a source of information on children's task solving processes. Combined with a rule-based, theory-driven automated scoring system, it could provide educators and educational psychologists with easy to obtain, in-depth information on children's task solving behavior, with a minimum of time and effort that needs to be invested by an examiner. This might enable process-oriented dynamic testing in the classroom, without any severe time investment of educators or psychologists, an often-voiced objection to the widespread use of dynamic testing (Sternberg & Grigorenko, 2002).

Analysis of process measures. For the analysis of the processes children use to solve certain tasks, Siegler (1987) pointed out that averaging data over multiple items could lead researchers to overlook aspects of the task solving processes. Investigating these task solving processes on the level of individual items (*Chapter 2 and 3*) was expected to provide valuable information regarding the potential factors contributing to children's success or failure when tested. The results underlined the importance of task characteristics in interpreting children's task solving processes. Previous research showed that the interpretation of completion time is dependent on task characteristics such as item difficulty (e.g., Dodonova & Dodonov, 2013; Goldhammer et al., 2014; Scherer, Greiff, & Hautamäki, 2015), and in line with this research, item characteristics were found to be an important explanatory factor of item success. In analyzing these complex and non-linear relationships, decision tree analyses provided a useful tool (Ritschard, 2014), and provided additional insights when combined with traditional, linear analyses.

Limitations and future research

Our findings should be viewed in the light of a number of limitations of the studies described in this thesis. Firstly, participants in all studies were primary school children of a single age range (7-to-8 years of age). More research might be needed to apply our findings to a more diverse population in terms of age, but also for the application in special education contexts.

The assessment of children in special education requires specific expertise, and it cannot be readily assumed that the processes that are used to solve the tasks are the same for children in special education contexts (Hessels, Vanderlinden, & Rojas, 2011).

Additionally, despite using TUIs to obtain non-obtrusive measurement of children's problem solving processes, children were also required to provide verbal explanations of their task solution. Requiring children to verbally explain their answers, may have influenced their task solving processes (Kirk & Ashcraft, 2001; Tenison, Fincham, & Anderson, 2014). To control for this, future research might employ a control group that is not required to verbally explain their answer, to provide more insight into the influence of the requirement of verbal explanations on children's task solving processes.

Greiff, Wüstenberg, and Avvisati (2015) discerned a number of factors that might prevent widespread use of log-file analysis, such as the complicated relationships between data, difficulty to determine the exact meaning of data patterns and their implications, the lack of technical expertise, and the fact the theory-driven research has never gained momentum. The complicated relationships between data, such as interactive relationships with task characteristics (e.g., Goldhammer et al., 2014), or curvilinear relationships between process measures and accuracy (e.g., Greiff, Niepel, Scherer, & Martin, 2016) could distort our understanding of the processes children use to solve tasks. To overcome this, future research might not only focus on linear relationship, but use more diverse methods to support our understanding of task solving processes, such as Rasch modeling (Stevenson, Hickendorff, et al., 2013), SEM modeling (Greiff et al., 2016), or decision tree analysis using C(A)RT and CHAID methods (McArdle, 2014; Ritschard, 2014). Additionally, Siegler (1987), recommended to look at the contributions to item success and be mindful of the dangers of averaging data over multiple items.

To extract exact meaning from the process measures, and allow for their application in recommendations towards educational interventions, future research might investigate the connections between task solving processes children use, and interventions that can be used to remediate processes that are not used efficiently. Through combining the processes within a specific domain of testing, with a specific intervention that targets the most relevant processes, assessment and intervention can be connected from an empirical basis, and provide a solid foundation for application in education. Future research might focus on the development and use of fully computerized dynamic tests, for instance through computerized versions of the Rey-Osterrieth complex figure, which would allow for automatized scoring of both accuracy and process measures. Through working with technology independently, children's learning potential and task solving processes can be uncovered, whilst decreasing the load for teachers. Using the technological opportunities that the 21st century has brought us might prove challenging, but may also provide valuable tools for the assessment and development of individual potential.