

## Content

Preface	
<b>ERIC KLOPP, JOHANN F. SCHNEIDER, ROBIN STARK</b> .....	<b>7</b>
Thinking Aloud: Foundations, Prospects and Practical Challenges	
<b>DIETER WALLACH, SEBASTIAN C. SCHOLZ</b> .....	<b>9</b>
Implementation of the Think Aloud Paradigm in a Study Analyzing the Solution Process of Figural Matrices	
<b>ALEXANDER KIRMSSE, SILKE ALINE RÖVER, NICOLAS BECKER</b> .....	<b>29</b>
Verbal Mediation of Planning in Elementary Students: A Think Aloud Approach	
<b>CATHERINE GUNZENHAUSER, FRANZISKA HAUCH, FRANZISKA STEPHAN, HENRIK SAALBACH</b> .....	<b>55</b>
Think Aloud Protocols in Medical Education: An Assessment of the Diagnostic Approach and Handling of Instructional Errors in Experts and Novices	
<b>MARTIN KLEIN, VERONIKA KOPP, MARTIN FISCHER, ROBIN STARK</b> .....	<b>71</b>
Cued Retrospective Recall Method Reveals Processes of Multimedia Learning	
<b>LISA STARK, ROLAND BRÜNKEN, BABETTE PARK</b> .....	<b>105</b>
The Three-Step Test-Interview (TSTI) for Pre-Testing Self-Completion Questionnaires	
<b>KEES VAN DER VEER</b> .....	<b>129</b>
Reflections on the Validity of the Think Aloud Method – the Case of Reactivity and a Possible Remedy	
<b>ERIC KLOPP, JOHANN F. SCHNEIDER, ROBIN STARK</b> .....	<b>157</b>
Authors .....	<b>167</b>

## Preface

This book is dedicated to the topic of generating data using the think aloud method. It updates and extends a collection of German papers published more than a decade ago (Schneider, 2001). This predecessor to the present book focused on the cognitive processes involved when respondents answer items of personality questionnaires. It collected works by researchers from the Psychology Department of Saarland University. The major goal back then was to provide an introduction to the think aloud method and to exemplify how the method can be used in the evaluation of questionnaire items.

The present book updates the former collection in three important aspects. Firstly, we chose English as a language in order to make our book available to international readers who are interested in applying the think aloud method in their own research. Secondly, the new book covers a broader range of topics. Instead of focusing on the application of the think aloud method in questionnaire construction, it now covers applications in such diverse fields as intelligence testing, developmental and educational psychology, and medical education. Furthermore, this broadening of content allowed us to include the work of researchers from several different universities. Lastly, this new book also reflects a change in theoretical reasoning about the think aloud method. Whereas initially, the think aloud method was mainly used in a descriptive and explorative manner, e.g., in problem-solving research to discover procedural aspects of cognitive processes, it is now used increasingly in a hypothesis testing manner.

In the first chapter, Dieter Wallach and Christian Scholz provide a precise and detailed introduction to the think aloud method. The first chapter is a comprehensive introduction into the theory behind the think aloud method as well as an overview of ways to use the method in applied settings.

The second chapter, written by Alexander Kirmsse, Silke Röver, and Nicolas Becker provides a close description of the implementation of the think aloud method in the validation process of intelligence test items. One remarkable feature of this contribution is that the think aloud method is used in a hypothesis testing way.

Catherine Gunzenhauser, Franziska Hauch, Franziska Stephan, and Henrik Saalbach provide an example from developmental psychology in the third chapter. Their paper is concerned with the cognitive processes children show while solving the Tower of London task. Once again, the think aloud method is used in a hypothesis testing way.

In the next chapter, Martin Klein, Veronika Kopp, Martin Fischer and Robin Stark deviate from the domain of psychology by taking the reader into the domain of medical education. The paper is framed within the context of medical diagnostics competences and provides an application of the think aloud method to describe how novices and experts deal with errors in the diagnostic process.

Lisa Stark, Roland Brünken, and Babette Park provide another variety of the think aloud method in educational psychology. Their paper covers the domain of multimedia learning. In contrast to the contributions before, the authors use the cued retrospective think aloud method.

The following chapter deals with the application of the think aloud method in questionnaire development. Kees van der Veer describes the so-called Three-Step Test-Interview. In this pre-testing practice, the think aloud method is used to improve the validity of questionnaire items.

The book ends with a short essay written by Eric Klopp, Johann F. Schneider, and Robin Stark about the validity of the think aloud method. The authors discuss the problem of reactivity and offer an idea for a novel approach to detect reactivity in think aloud protocols using meta-comments.

It is our hope that the present collection of papers encourages intensive application and further development of the think aloud method for the study of cognitive processes in applied research settings.

Last but not least, the editors want to thank PD Dr. Christoph Paulus for his help in reviewing the first versions of the chapters.

*Eric Klopp*

*Johann F. Schneider*

*Robin Stark*

#### **References**

SCHNEIDER, J. F. (2001) (Eds.) *Lautes Denken – Prozessanalysen bei Selbst- und Fremdeinschätzungen* [Thinking aloud – process analyses on self- and external assessments.] Weimar: Daddler.

## Thinking Aloud: Foundations, Prospects and Practical Challenges

Dieter Wallach, Sebastian C. Scholz

Since the beginning of the 20th century, the *thinking aloud method* is – despite of continuing controversial discussions – one of the most commonly used methods to gather qualitative evidence about mental processes (Sauro, 2016). Bühler (1907) applied the method in his analyses of cognitive processes during problem solving activities. Duncker (1935) and Claparède (1971) used thinking aloud data to study the course of human thinking processes, De Groot (1965) resorted on the thinking aloud method to investigate mental processes of chess players. After some quiescence in the bloom of Behaviourism, thinking aloud and the use of verbal data experienced a renaissance (see the seminal work of Newell and Simon, 1972) in the so-called *Cognitive Revolution* of the late 1960s. In the German psychological literature, Lüer (1973) and Dörner (1981) demonstrated the benefits of using thinking aloud in their investigations of complex cognitive processes. More recently, the thinking aloud method is extensively used in applied research on *Human-Computer Interaction*. In usability testing settings participants are encouraged to think aloud while working with interactive systems. Jacob Nielsen (2012) ennobles the method as the “number one usability tool” that “serves as a window to the soul” of participants “letting you discover what users really think about your design” (see also Dumas & Loring, 2008).

In thinking aloud studies, participants are asked to verbally express their thoughts while working on a task. The fundamental underlying hypothesis of thinking aloud is that the content of working memory is verbally conveyed and that the resulting qualitative data can be used to infer the knowledge involved in dealing with the task. Strictly speaking, the term “*thinking aloud*” is not entirely correct, as collateral to the products of thinking, a participant’s perceptions, emotions, and thoughts on her own thoughts and actions (meta-cognition) are reported. In contrast to *Introspection* (Wundt, 1911; see Sauro, 2016), participants are not explicitly encouraged to reflect on their own thinking processes. Instead,

## **Implementation of the Think Aloud Paradigm in a Study Analyzing the Solution Process of Figural Matrices**

Alexander Kirmsse, Silke Aline Röver, Nicolas Becker

With this chapter, we want to provide insights into the way we conducted a study to assess qualitative aspects of the solution process of figural matrices by using the think aloud method. Since theoretical considerations related to the solution process formed the foundation for the study, they are presented in the first section. The second section deals with the research goal of the study that influenced the implementation of the think aloud method. In the third and fourth section, we provide information on the model based derivation of the coding scheme, the acquisition of the verbal material and its coding as well as on the determination of the inter-rater reliability. We also shortly describe the results obtained by the think aloud method and illustrate how these findings contributed to the understanding of the solution process of figural matrices. Finally, we discuss aspects of the way the study was conducted and evaluate methodological issues of the think aloud paradigm.

### **Components of Figural Matrices**

Figural matrices are one of the most common item formats used for the assessment of intelligence (Jensen, 1998). Figure 1 shows an example of a classical figural matrix item. The item stem consists of a  $3 \times 3$  matrix of cells with geometrical symbols (e.g. figural elements). These elements follow certain design rules. In the example, the arrow is rotated by  $90^\circ$  in a counterclockwise direction throughout the rows of the matrix. The circles follow an addition rule over the rows, that is the circles appearing in the first and the second cell of a row add up in the third cell. The last cell of the item stem (solution cell) is left empty. In the lower part of Figure 1, the response format can be found. It consists of the correct solution (attractor) which

## **Verbal Mediation of Planning in Elementary Students: A Think Aloud Approach**

Catherine Gunzenhauser, Franziska Hauch, Franziska Stephan, Henrik Saalbach

Children and adults often use verbal strategies when working on challenging tasks (for summaries see Alderson-Day & Fernyhough, 2015; Winsler, 2009). In younger children, this *self-directed speech* frequently takes the form of audible self-talk, while children in elementary school as well as adolescents and adults tend to use verbal thinking that is not accessible to observers. In consequence, investigating self-directed speech in elementary school and beyond is a methodological challenge. A number of methods have been applied to investigate that phenomenon (see Alderson-Day and Fernyhough, 2015), among others, self-reports in the form of retrospective questionnaires and experience sampling, neuroimaging, or dual-task paradigms which try to estimate effects of internal self-directed speech by creating interferences. However, these methods do not allow a direct investigation of the contents of self-directed speech. In the present chapter, we describe a think aloud approach to studying the verbal mediation of planning in elementary students. Building on findings on the quantity of self-directed speech in younger children, we focused on origins and consequences of individual differences in self-directed speech. Moreover, we aimed at examining whether a higher quality of self-directed speech might prove beneficial for elementary students' planning success.

### **Development of self-directed speech**

In early and middle childhood, children often talk to themselves aloud while they are engaged in an activity. This *self-talk* or *private speech* peaks during the preschool years (for a summary see Winsler, 2009). After age five, children's private speech becomes less overt, being gradually transformed into *inner speech* with some external manifestations such as muttering, whispering or lip movements (Winsler & Naglieri, 2003). In children after age nine and adults, such external

## **Think Aloud Protocols in Medical Education: An Assessment of the Diagnostic Approach and Handling of Instructional Errors in Experts and Novices**

Martin Klein, Veronika Kopp, Martin Fischer, Robin Stark

Diagnostic errors in medical practice often have dramatic consequences (Al-Assaf, Bumpus, Carter, & Dixon, 2003) such as medical mistreatments and increased patient mortality (Hashem, Chi & Friedmann, 2003; Plebani, 2009; Croskerry, 2009). Despite advances in medicine technology and programs aiming at improving the diagnostic process (e.g. Mamede, Schmitt, & Rikers, 2007; Lawson & Daniel, 2011; Croskerry, 2009), still 10-15 % percent of all diagnoses are estimated to be wrong (Berner & Graber, 2008; Croskerry, 2009; 2013). Finding the right diagnosis seems not only a difficult task for medicine students (e.g. Gräsel & Mandl, 1993), but for experienced practitioners as well.

Facing these problems early on by fostering diagnostic competence during studies at university is therefore an important objective in medical education. In the present study, diagnostic competence was conceptualised as a combination of *conceptual knowledge* that consists of topic-specific declarative knowledge about concepts and relations, of *strategic knowledge* about diagnostic problem-solving strategies and heuristics, and of *conditional knowledge* about the justification of diagnostic procedures and decisions (Paris, Lipson, & Wixson, 1983; Stark, Kopp, & Fischer, 2011). Multiple studies have implemented a variety of different didactic approaches such as case-based learning or worked examples (Renkl, 2005) to successfully foster diagnostic competence (e.g. Kopp, Stark, Kühne-Eversmann, & Fischer, 2009; Koca, Hege, Kopp, & Fischer, 2008; Fischer, Aulinger, & Kopp, 2005). Worked examples, however, have been shown to lose their effectiveness with increasing expertise (Renkl, 2005; cf. *expertise reversal effect*, Kalyuga, Renkl, & Paas, 2010), which endangers the purpose of reducing diagnostic errors over the course of medical training.

## **Cued Retrospective Recall Method Reveals Processes of Multimedia Learning**

Lisa Stark, Roland Brünken, Babette Park

Multimedia theories propose different cognitive processes, which build the basis for successful learning with multimedia instruction (e.g., Mayer, 2014). However, these processes are not supported empirically as they have not yet been the explicit focus of research. By applying different measures for learning performance (e.g., retention and transfer tests), learning processes have only been addressed implicitly until now. Furthermore, within the last decades, multimedia research has started to integrate further motivational or affective variables (Park, Plass, & Brünken, 2014) and has begun to include online measures of learning processes, e.g., eye tracking (e.g., Jarodzka, Scheiter, Gerjets, & Van Gog, 2010; Park, Knörzer, Plass, & Brünken, 2015; Van Gog et al., 2005). However, eye tracking measures do not necessarily reveal information about learners' cognitions or the processes of knowledge acquisition. Although, the *eye-mind hypothesis* (Rayner, 1998) postulates that information is processed while it is watched, one can easily imagine situations in which one looks at something while thinking of something else. Therefore, further methods are required in order to catch the learners' cognitions. Think aloud protocols (Ericsson & Simon, 1993) are often used but they also face some methodical problems, e.g., reactivity with task performance (see below). Thus, a combined method of eye tracking data and think aloud method, namely the *cued retrospective think aloud* method, seems to be promising to provide data on learning processes. In the present study, this method is applied in order to investigate processes during learning with multimedia instruction.

Processes in multimedia learning are investigated applying the cued retrospective think aloud method. Verbalized cognitions are analyzed using a category system summarizing different multimedia processes. Afterwards, high and low performing learners are compared regarding verbalized multimedia processes.



## **The Three-Step Test-Interview (TSTI) for Pre-Testing Self-Completion Questionnaires\***

Kees van der Veer

\* The report of the development, description and application of the TSTI in this chapter is largely based upon Hak, T., K. Van der Veer, & H. Jansen (2008). The Three- Step Test-Interview (TSTI): An observation based method for pretesting self-completion questionnaires. *Survey Research Methods*, 2(3), 143-150.

Collecting data during a survey requires questionnaires that are able to generate reliable and valid data. That means in the first place that the wording of questions, whether open or closed, requires thorough pretesting in some way in order to reduce possible errors. Every handbook on questionnaire design includes sets of criteria for developing adequate questions (see e.g. Sayeed, 2015). Pretesting is often done or at least starts with desk expert reviewing of questions, followed by a pilot, irrespective of the mode of questionnaire administration (interview of self-completion, and whether presented by mail, email, or otherwise).

Increasingly and in addition, non-sampling data error in surveys is analyzed as resulting from problems that might occur in the response process, i.e. the process of interaction between the instrument (questionnaire) and the respondent, irrespective of the mode (paper and pencil, telephone, internet) by which the respondent is approached.

When a respondent responds to a questionnaire, a problem may arise at any step in the response process (as e.g. described by Tourangeau 1984) at any point in the completion of the questionnaire. When such a problem occurs, data error might or might not result. Cognitive interviewing has been developed as an instrument for identifying such problems in the response process, their localization (both in the response process model and in the questionnaire), their effects (in terms of data error), and their causes.

In current pre-testing practice the term 'cognitive interviewing' refers to two main techniques, think aloud and probing (see, e.g., Willis, 2005; Willis, Schechter, & Whitaker, 1999). These two techniques are very different in terms of their aims and of their methodological status. Think aloud was developed and

## Reflections on the Validity of the Think Aloud Method – the Case of Reactivity and a Possible Remedy

Eric Klopp, Johann F. Schneider, Robin Stark

The think aloud method is used with increasing frequency in psychology to capture cognitive processes, and is nowadays widely accepted in applied settings (see Fox, Ericsson & Best, 2011; Güss, 2018; Sudman, Bradburn & Schwarz, 1996; Pepper, Hodgen, Lamesoo, Kõiv & Tolboon, 2016). However, critical aspects of the think aloud method regarding its validity have also been discussed (see Sasaki, 2008). The main critique of the think aloud method concerns validity aspects and refers to the so-called problem of reactivity. Reactivity means that “verbalizing changes the process *because* of the verbal report: the process studied can be temporally extended, disturbed or otherwise manipulated” (Wallach & Scholz, this volume). In this essay, we reflect on the issue of reactive think aloud protocols drawing on recent studies. Our aim is to present a potential remedy by suggesting a method to detect reactivity in think aloud protocols and derive a research strategy that allows to control possible effects of reactivity.

Following Stratman and Hamp-Lyons (1994, p. 95), there are five conditions that potentially cause reactivity in think aloud protocols. The first condition is met when the experimental task prompts subjects to elicit an inappropriate level of verbalization. The second condition is met when participants have a limited short-term memory capacity for speaking and concentrating at the same time. The third condition applies when hearing their own voices distracts participants in a think aloud study. The fourth condition refers to learning that occurs when thinking aloud increases subjects’ critical attention to their activities. The fifth and last condition applies when the experimenter directly or indirectly influences the participant through verbal and nonverbal cues. These conditions, stated by Stratman and Hamp-Lyons (1994), can be classified into two overarching types. The first type (hereafter: type I) comprises conditions one and five. The communality of these conditions is that they refer to the (experimental) setting in which the think aloud method is applied. The second type (hereafter: type II) subsumes the remaining conditions two, three, and four. All of these conditions refer to characteristics of the participants.