

Annales Universitatis Paedagogicae Cracoviensis

Studia Anglica I (2011)

Anna Konieczna

MIRRORING THE WORK OF MIND: VERBAL REPORT AS REFLECTING THE USE OF STRATEGIES

Introduction

The text presented here may be considered a continuation of the discussion that so far has evolved under several titles of mine (Konieczna 2009a, 2009b, 2011). The discussion concerns what can be denoted as theoretical foundations underlying the use of the think aloud (TA) variety of verbal protocol in the research on reading. The paper explores the metaphor of a mirror to discuss the validity of strategies reported through verbal reports. It should be admitted that the mirror metaphor is a very attractive one, as it suggests providing a comfortable way of doing observation on mental work. Sainsbury (2003: 134) used a similar metaphor of “a window into thoughts,” also emphasizing the supposedly easy access that the think aloud verbal report offers to mental phenomena. Elsewhere the author explains that the procedure is likely to “capture some of the ephemeral processes of understanding” (Sainsbury 2003: 131). The paper presented here, though, tries to indicate that the “mirror of TA,” depending on how it is used, may in fact reflect different phenomena. Many times it may be a distorting mirror, or at least one with scratchy surface, reflecting outlined images but making the details indiscernible, vague, and uncertain. All the comments included in this paper base on what researchers using the TA procedure report and on the author’s own experience with the use of that method.

Retrospective and concurrent verbalizations

The optimism and approval surrounding the TA procedure, reflected also in the above mentioned metaphors, appears to result mainly from the disappointment with retrospective methods of cognitive research. Retrospective collection of data involves the retrieval of information from memory. It is realized after the primary task has been completed and requires the respondents to report what they remember doing or thinking. Yet, due to the consequences of time delay between task performance and reporting, retrospection is believed to be exposed to different sources of non-validity. For example, when reporting processes, respondents may not stick to one particular instance of performance, but may instead refer to their general knowledge of the tasks and how they deal with them generally. Respondents who are aware of what they normally do in similar tasks may be likely to report this,

without taking recourse to any particular situation (Ericsson and Simon 1993: 23). Retrospection is also very likely to misrepresent the sequence of steps taken while doing a task, as the information heeded by the respondent in particular moments of time can be easily confused with information attended to subsequently (Ericsson and Simon 1993: 19). One other point is that the retrieval of information attended to in the past “is an onerous task” (Ericsson and Simon 1993: 20), thus many respondents will find it easier to generate the information instead, which means that they will not really report on what happened, but will elaborate on the processes supposedly used only at the time of providing the protocol. “It is reasonable to assume that the subject either infers general motives or processes from retrieved selected episodic memories, or tries to rationalize his behavior using other sources of information than the memory of the processes” (Ericsson and Simon 1993: 46). Ericsson and Simon point to the fact that this ad hoc inference is very likely to take place while responding to the ‘why’ kind of questions. As the authors explain, answers to ‘why’ questions are often generated a posteriori, only at the time of asking, and are not based on the memory trace but inferred. Moreover, when retrospecting, subjects may draw from their knowledge and beliefs on what should be done in certain tasks (Ericsson and Simon 1993: 23; Garner, Wagoner and Smith 1983: 440). Research on retrospection has shown that subjects often “report using behaviours they do not demonstrate using” (Garner, Wagoner and Smith 1983: 440), fabricate strategies at the time of speaking and “fail to report obstacles or resolutions to obstacles apparently deemed too obvious to mention” (Garner, Wagoner and Smith 1983: 440). The limitations of retrospection have been mentioned by various researchers as rationalizing the use of the supposedly more veridical method of cognitive research which is the introspective think aloud protocol. Think aloud was believed to offer a way out of the memory problems described.

Think aloud is a kind of verbal report realized concurrently with the main task performed (called *primary task*). The procedure asks respondents to vocalize thoughts while performing the task. The assumption underlying the use of TA is that the sequence of the information heeded while performing a task is accessible and can be represented through reporting continually the content of short term memory. The TA procedure requires the researcher to draw inferences concerning the strategies taking place – the respondent is not encouraged to theorize on what he or she is doing, as this is considered to disturb performance.

The most thorough work on verbal reporting that has been referred to most often is the one by Simon and Ericsson (1984, 1993). Mere reference to these authors is often considered as validating the methodology used and as pointing to theoretical foundations underlying the methodology. The point I would like to make in this section, though, is that the reference made in some research reports is quite superficial, in the sense that it relies on mentioning the names of the authors but is not followed by strict adherence to the criteria the authors establish. This introduces much confusion to the research done with the use of TA, as the procedures which various authors use tend to differ considerably, in spite of the superficial similarity of the terminology used. This paper will not discuss the theory of Ericsson and Simon but will point to the methodological discrepancies in the way the theory is put into practice, and, taking the stance of Ericsson and Simon’s perspective, will point to the dangers of non-validity introduced by the differences in TA usage.

Ericsson and Simon distinguish three levels of verbalizations in verbal reports, stating, at the same time, that only level 1, consisting of the “vocalization of covert articulatory or oral encodings” and level 2, involving additional recoding of non-verbal information into verbal code, can produce a protocol which is valid and which does not alter the structure of the cognitive processes taking place while doing a particular task. Verbalizations denoted as belonging to the third level, and involving not only articulation but also explanation of thoughts, ideas and motifs, filtering/ selecting information to be reported, or reporting on subconscious or automatized information will change the structure of the thought processes and thus invalidate the protocol (Ericsson and Simon 1993: 18–19, 79–80). Think Aloud is supposed to exemplify level 2 verbalization, displaying just the thoughts as they appear and thus not altering the structure of thought processes in which the subject is involved. TA protocols “reflect states of heeded information and do not describe the details of the information nor why that particular information was heeded” (Ericsson and Simon 1993: xxxv), thus “only in the context of a task analysis can one make sound inferences about the sequence of underlying cognitive processes.” (Ericsson and Simon 1993: xxxv). Even though the TA verbalization is very often fragmentary, full of false starts and stops in the middle of the sentence, apparently lacking organization, researchers need to ‘resist’ the urge to elicit coherent and logically complete verbalizations. Descriptions and explanations in the protocols would add coherence to the verbalization, however, they would deviate the processes used for task performance (Ericsson and Simon 1993: xv). It is not the respondent’s, but the researcher’s work to infer cognitive processes from utterances which often apparently lack coherence and are fragmentary, and which do not answer the ‘why’ question. At the same time, it is very important for the appropriate use of the TA procedure that respondents are warned not to explain, describe, analyse or interpret their own thought processes (Ericsson and Simon 1993: xiii). All the interpretation, theorizing and inference belongs to the researcher.

What draws attention is the fact that some research done with what comes to be called *TA* does not fulfill the requirements of Ericsson and Simon’s level 2 verbalization, and thus, at least according to the original theory, is likely to be reactive (changing the main process). Various researchers, as if ignoring Ericsson and Simon’s statements and at the same time contradicting the essence of think aloud, ask respondents for reasons, descriptions and explanations. Jimenez (1997: 233) writes that “the think aloud procedure consists of [...] asking the participant to describe and explain [...] what he or she is thinking about.” Scott (2008: 302–303, 316) in her TA coding scheme includes the category of *verbalizing actions taken to process the text*, giving as an example the following verbalization: “I was going to skip and then try to find it again try to figure it out... [sic].” This verbalization is clearly a description of actions, which should be avoided within the TA methodology. With reference to the possible kinds of verbalizations elicited, Ericsson and Simon also recommend the use of the “keep talking” prompt for situations in which the respondent lapses into silence. Prompts of the “what are you thinking about?” kind should be avoided as answering them “requires a description of the thoughts during the silence in a way that may disrupt the thought sequence” (Ericsson and Simon 1993: xxviii). In spite of that, researchers often admit to be using the “what are you

thinking about?” or similar prompt (Scott 2008: 300; Braten and Stromso 2003: 203), or the one that seems even more likely to elicit description instead of report: “What do/did you think about this/that?” (Jimenez, Garcia and Pearson 1996: 97; Scott 2008: 300). What is more, sometimes researchers decide to use probes – questions of a specific kind concerning particular information (see for example Jimenez, Garcia and Pearson 1996: 97). Probes, though, foster the selection of information and attending to information which otherwise might not have been attended to all. Thus they clearly elicit what was called level 3 verbalization and what was described as possibly modifying the sequence of thoughts. Respondents may infer or generate the information according to the probe, especially if they notice that is what the experimenter is interested in or what he wants to hear. They may act in compliance with these perceived expectations, and thus bias the data provided.

The here presented examples of unexplained divergence from what can be considered original and theoretically grounded form of the TA procedure is the first “scratch on the mirror.” This divergence may be believed, at least within the framework of Simon and Ericsson’s theory, to distort the image the metaphorical TA “mirror” reflects.

Overt and covert processes

The distinction made by O’Malley and Chamot between overt and covert strategies can be taken as a starting point in the discussion on what TA is and is not likely to reflect. O’Malley and Chamot (1990: 87–88) define overt strategies (like using a dictionary or note taking) as those which can be easily observed, and covert strategies as requiring “introspective forms of data collection in which the informant provides a description of the strategy used.” At the same time the authors admit that “strategies that occur overtly cannot qualify as mental processes” (O’Malley and Chamot 1990: 88).

The analogy I would like to draw between the distinction introduced by O’Malley and Chamot and verbal report data is that in the think aloud procedure some of the strategies are much easier noticeable than others (even though they are not completely ‘overt’ in O’Malley and Chamot’s sense). Still, they are easy to notice as in fact they do not require inference on the researcher’s part. They simply appear and can be observed while the respondent verbalizes. Questionable, though, is the extent to which the strategies can be considered as offering insight into people’s mental processes and whether at all they should be defined as mental. In order to discuss the tension between what the TA methodology claims to be doing (offering insight into mental phenomena) and what it appears to be doing, at least within some of its applications, I will quote Green’s data collected on a reading comprehension task. Green codes her TA data as follows. For student 1: read/re-read/read/re-read/read/focus/retrieve/read/read question 1/re-read question 1/re-read/re-read question 1/scan/read question 2/re-read question 1/respond. For student 2: read/focus/re-read/read/focus/re-read/read question 1/focus/keywords search/respond (Green 2009: 71–72). As can be noticed, the author coded most frequently and mainly instances of reading and re-reading. Noticing the respondent read while he or she is asked to verbalize is an observation quite easy to make. It does not

require inference on the researcher's part and does not reveal much about the way the text is comprehended. The final of the strategies coded as *responding* could be in fact considered entirely overt, thus not mental. The conclusions Green makes basing on the coding are the following:

We begin by considering Student 1. From the coded segments, this student appears to go through a cycle of reading followed by re-reading sections of the passage. At segment 6 the student 'focuses', perhaps on a particular word or phrase, and then retrieves the meaning of that particular word or phrase. This could suggest an initial failure to understand and then retrieval of the appropriate meaning. The student then reads the question, re-reads the questions and then re-reads some text. Further cycles of re-reading the question and the text then ensue before the question is answered. Student 2 approaches the task differently. Student 2 reads a section of text, focuses on a phrase or word and then continues to read. The cycle is repeated until the passage is read. Responding to the first question of the text, the student reads the question, focuses on a particular word or phrase and then searches the text for that word or phrase. The question is then answered. (Green 2009: 72)

The first comment that can be made with reference to the data collected by Green is that certain processes appear quite 'overt' within TA, and that they do not really denote mental processes, but rather physical actions – concrete, easily observable behaviours. The easiest thing to observe is the sequence of actions taken (see also the study of Cohen and Upton 2007: 221; Konieczna 2009a: 156–158) and the particular actions/behaviours (like reading, re-reading) themselves. Here cognitive processes still stay hidden or are only minimally inferred. What is mainly coded is just the approach taken (approach externally observable) to complete the task. The TA procedure realized in this way is by no way 'a mirror' of cognitive processes; it can only be considered a mirror of some externally observable actions. Moreover, we may argue, as I have done elsewhere (Konieczna 2011), that some of the strategies most frequently coded in the TA on reading, for example *reading*, *re-reading*, *paraphrasing*, *translating* or *summarizing*, may be partly induced by the characteristics of the TA procedure as such (which asks respondents to report on the text read). Thus conclusions concerning their use for the sake of text comprehension may be biased. The overwhelmingly frequent coding of potentially TA-induced phenomena is pertinent not only to Green's study, but reappears in various reports (see for example Cohen and Upton 2007) including my own ones (Konieczna 2009a). Researchers (see Cohen and Upton 2007: 222) sometimes admit having developed an elaborate framework of coding rubrics before the actual coding took place, and having to reduce the scheme to just several chosen categories at the time of coding, which clearly suggests a kind of 'unexpected narrowness' of the quality of the data obtained. Researchers as if expect more than they really appear to get in the end. Normally the categories developed a priori need to be recombined, with some categories melded, and some skipped due to the scarcity of their appearances in the protocols. The difficulty here described, which is a first hand experience of mine, points to the fact that what the "mirror" of TA on reading reflects, may constitute a quite narrow range of the phenomena expected.

Misperformance within TA

The next dilemma that relates to the phenomena likely to be reflected with the use of TA concerns respondents' misperformance. More specifically, the problem concerns the classifications of mistaken trials to do something. I must admit I was faced with the dilemma of how to classify mistaken performance of respondents, and I finally decided to treat particular behaviours as 'trials' to do something, ignoring the extent to which the trial was successful. On the other hand, Green (2008: 82) chooses to use separate codes for *mis-reading text* and *mis-interpreting text*. Mistaken performance is also coded in the studies of Scott (2008: 303) and in the report of Kendeou, Muis and Fulton (2011: 6). The trouble that appears here, though, consists of the fact that what is being done is simultaneous coding and assessment, while not providing any kind of explicit assessment criteria. My own research shows that the seriousness of mistaken performance will most often be quite different. Sometimes, like for example in case of *interpretation*, or even more often in case of the strategy denoted in my research as *identifying the main idea*, I experienced serious problems concerned with deciding to what extent the idea given by the student could be considered correct or incorrect. Very often the thing was almost true, and had just some shades of incorrect interpretation. Even more often the trouble was that the main idea had been formulated in a way which I considered not quite complete or too general, and thus not fully true about the fragment considered. My opinion is that assessing performance without explicitly formulated rating scales may bias the results. In my research, as was noticed earlier, I decided to abandon the distinction into what was correct and what was incorrect. After all, a strategy may be used successfully or not, but it still keeps being the same strategy. Assessing the result of its use is quite a different thing and requires additional criteria. Most authors working with TA do not quote mistaken productions, nor even comment on the issue, basing on which we can assume that they code all instances of strategies use, not taking into account whether these are successful or not. Still, there are also cases in which authors try to deal with the issue of correctness within the TA methodology. Lau (2006), investigating good and poor readers of Chinese, introduces into the coding system differentiated weighting of strategies, depending on how successful they are (Lau 2006: 388). Using this rating Lau calculated respondents' strategy use score. Yet, such a rating relates not merely to strategy use, but rather to the successfulness of strategy use, which might be a slightly different issue.

As the issue of mistaken performance has so far remained, to a large extent, unresolved, its differentiated treatment within various research studies may influence the data coded. This, metaphorically, may add to the scratchy surface of the TA "mirror."

Conclusion

The observations made in the paper point to the fact that the "access to mind" that the think aloud verbal report offers is much more problematical than it seems. The Think Aloud "mirror," which is supposed to reflect the work of mind, may misrepresent certain facts. This misrepresentation is even more likely to take place

if researchers are not very specific about the procedure they use or unreflectively modify the procedure.

Bibliography

- Braten, I., Stromso, H.I. 2003. "A Longitudinal Think-Aloud Study of Spontaneous Strategic Processing During the Reading of Multiple Expository Texts" in *Reading and Writing: An Interdisciplinary Journal*. 16 (3): 195–218.
- Cohen, A.D., Upton, T.A. 2007. "I want to go back to the text': Response strategies on the reading subtest of the new TOEFL" in *Language Testing*. 24 (2): 209–250.
- Ericsson, K.A., Simon, H.A. 1993. *Protocol Analysis. Verbal Reports as Data*. Cambridge, MA–London, UK: A Bradford Book, The MIT Press.
- Garner, R., Wagoner, S., Smith, T. 1983. "Externalizing question-answering strategies of good and poor comprehenders" in *Reading Research Quarterly*. 18 (4): 439–447.
- Green, A. 2009. *Verbal Protocol Analysis in Language Testing Research*. Cambridge: CUP.
- Jimenez, R.T. 1997. "The Strategic Reading Abilities and Potential of Five Low-Literacy Latina/o Readers in Middle School" in *Reading Research Quarterly*. 32 (3): 224–243.
- Jimenez, R.T., Garcia, G.E., Pearson, P.D. 1996. "The Reading Strategies of Bilingual Latina/o Students Who Are Successful English Readers: Opportunities and Obstacles" in *Reading Research Quarterly*. 31 (1): 90–112.
- Kendeou, P., Muis, K.R., Fulton, S. 2011. "Reader and text factors in reading comprehension processes" in *Journal of Research in Reading*. 34: 1–19.
- Konieczna, A. 2009a. "The Use of Strategies in Gapped-Text Reading Task" in *Current Issues in English Studies*. Kraków: Wydawnictwo Naukowe Uniwersytetu Pedagogicznego.
- Konieczna, A. 2009b. "Raport werbalny i jego wykorzystanie w badaniach nad strategiami rozwiązywania testów z zakresu czytania" w *Neofilolog*. 33: 77–88.
- Konieczna, A. 2011. "Theoretical considerations concerning the use of Think Aloud protocol in the research on testing reading" in *Texte in Bewegung setzen. Aufsätze zur Literatur- und Fremdsprachendidaktik*, Dresden–Wrocław: Neisse Verlag.
- Lau, K.L. 2006. "Reading Strategy Use Between Chinese Good and Poor Readers: a Think-Aloud Study" in *Journal of Research in Reading*. 29 (4): 383–399.
- O'Malley, J.M., Chamot, A.U. 1990. *Learning Strategies in Second Language Acquisition*. Cambridge–New York–Port Chester–Melbourne–Sydney: CUP.
- Sainsbury, M. 2003. "Thinking aloud: children's interactions with text" in *Reading Literacy and Language*. 37 (3): 131–135.
- Scott, D.B. 2008. "Assessing Text Processing: A Comparison of Four Methods" in *Journal of Literacy Research*. 40 (3): 290–316.

Lustrzane odbicie pracy umysłu – raport werbalny jako odzwierciedlenie stosowanych strategii

Streszczenie

Artykuł dotyczy metodologii prowadzenia badań nad procesami poznawczymi z wykorzystaniem raportu werbalnego, ze szczególnym uwzględnieniem zyskującego na popularności protokołu głośnego myślenia. Przywołana w tytule artykułu metafora lustrzanego odbicia pracy umysłu doskonale obrazuje optymizm panujący wśród badaczy wykorzystujących protokoły

głośnego myślenia. Artykuł tu przedstawiony wskazuje jednak na fakt, że optymizm ten nie jest w pełni uzasadniony. Autorka ukazuje liczne ograniczenia opisywanej metody badawczej. Odnosi się do prac badawczych autorów wykorzystujących protokół głośnego myślenia i ukazuje niespójności metodologiczne w obrębie tych badań. Ograniczenia metody i problemy z wykorzystaniem protokołu głośnego myślenia, na które wskazuje artykuł, to m.in.: występujące pomiędzy poszczególnymi badaniami niezgodności w sposobie zastosowania metody, ignorowanie podstaw teoretycznych metody opisanych przez Ericssona i Simona (1993), wnioskowanie w dużej mierze o strategiach jawnych, które można z łatwością obserwować, niewielka ilość inferencji dotyczących niejawnych procesów umysłowych. Artykuł opisuje ponadto cechy charakterystyczne raportów werbalnych realizowanych w sposób retro- i introspekcyjny oraz rodzaje raportów według klasyfikacji Ericssona i Simona (1993).