

Abstract

This study evaluates the effectiveness of the think-aloud method used during cognitive interviewing for survey – item testing with children. This study examines the application of a think-aloud technique during testing of a computer-based assessment, across two different grades and age-groups: grade 4 (age 9-10) and grade 8 (age 14-15). We compared the effectiveness of the think-aloud method between the two grades by evaluating the verbal reports length, time of response, and the quality and the completeness of the student responses. We found that the think-aloud method does not vary in the quality of the verbal output obtained from children in the two age groups. Although research suggests that young children typically find the think-aloud method to be difficult (Someren et al, 1994), this research shows that there is no difference in the effectiveness of the think-aloud method between students in Grade 4 and in Grade 8.

Introduction

High quality survey data often requires laboratory pretesting of the survey questionnaires or educational assessments. During the last two decades, cognitive interviewing became one of the most widely used pretesting methods. Cognitive interviewing is intended to uncover problems with the questions that may compromise the quality of responses before survey or assessment fielding. One of the procedures used during the cognitive interviews is the think aloud that requires respondents to verbalize their thoughts as they answer survey or test questions.

There is little empirical research on the effectiveness of the think-aloud method utilization during the pretesting of survey or assessments with children. Previous research has shown mixed results. Some reported that young children and teenagers lacked the ability or motivation to articulate their thoughts, while others showed that the think aloud works well with teenagers but requires age relevant instructions and more follow-up probing with younger children (Leeuw, et al. 2004; Someren et al, 1994) .

Objectives

- To evaluate effectiveness of the think aloud procedure and quality of the verbal protocols produced by children in Grade 4 (ages 9-10) and Grade 8 (ages 14-15).
- To compare the effectiveness of think aloud procedure and quality of the verbal protocols between children in Grade 4 (ages 9-10) and Grade 8 (ages 14-15) .
- To compare ability to recognize and categorize information from verbal reports into four processes – comprehension, retrieval, judgment, and response – identified the Question Response Model (Tourangeau, 2000) across grades.
- To evaluate ability to evaluate question processing at each of four response stages (comprehension, retrieval, judgment, and response) based on the information provided in the student think-aloud.

- Ten 4th grade students and ten 8th grade students participated in cognitive interviews.
- Each student group completed 4 grade-specific science questions.
- Concurrent think-aloud procedure utilized during the cognitive interview.
- Students were provided with detailed think-aloud instructions, a demonstration, and a think-aloud practicing session before attempting to answer each question while thinking-aloud.
- Consistent think-aloud procedure and instructions across students ensured with detailed script.
- The following 4 factors were controlled for between 4th and 8th grade in terms of question type assignment:
 - Subject:** Science NCES Questions administered in 2009
 - Content Classifications:** Physical Science, Earth and Space Science, Life Science
 - Question Difficulty:** Easy, Medium, Hard
 - Question Order Presentation**

Demographics

- A total of 20 students completed the cognitive interview. The sample consisted of a mix of SES, race/ethnicity (Table 1) and gender (Table 2).

Table 1. Race/Ethnicity by Grade

	Grade 4	Grade 8	Total
Black or African American	2	3	5
Hispanic or Latino(a)	4	2	6
White or Caucasian	4	5	9
Asian	0	0	0
Other	0	0	0
Total	10	10	20

Table 2. Gender by Grade

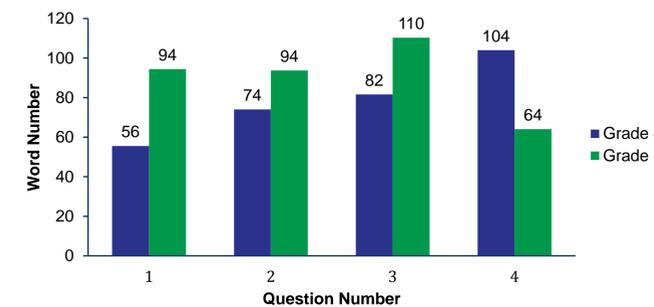
	Grade 4	Grade 8	Total
Female	4	8	12
Male	6	2	8

- Controlled for experience and frequency of use of technology (tablet, phone, and computer) across grade 4 and grade 8 students.

Results

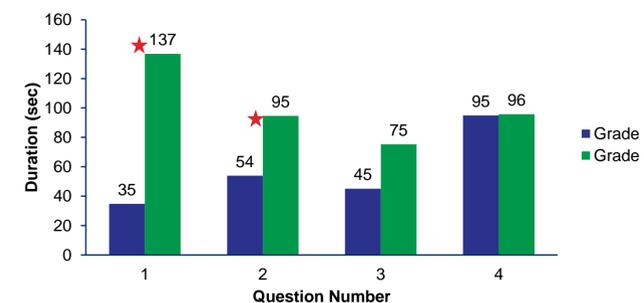
- There was no significant difference in the verbal reports length (number of words) provided during the student think-aloud across questions between grade 4 and grade 8 students.

Fig. 3. Average Number of Words in the Student Verbal Report by Grade



- We found a significant difference in the amount of time the student spent doing a think-aloud across student grades. Overall, grade 8 students took more time on their think-aloud than did grade 4 students.

Fig.4. Duration (sec) of the Student Think-Aloud by Grade

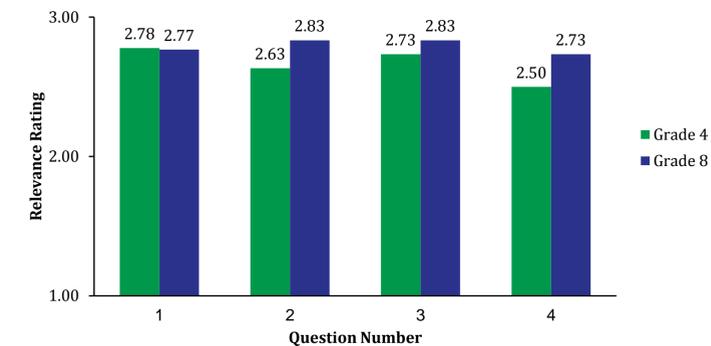


Results

Relevance of the Verbal Reports

- There was no significant difference in the relevance of the student think-aloud to the topic of the question between the two grades.
- Both 4th grade and 8th grade student’s think-aloud remained equally focused on the topic of the question.

Fig. 5. Relevance of the Verbal Reports to the Question Topic by Grade



Response Process Evaluation

- We examined whether there are differences in the ability to evaluate question processing at each of four response stages of the question response process (3-point Likert Scale: 1 – Unable, 2 – Somewhat, 3 – Able).
- Overall, we found no significant differences between grades in the ability to evaluate information at each of the response stages.
- Results showed differences in ability to identify and evaluate the judgment process between the two grades, where it was more difficult to assess judgment in 4th graders than in 8th graders.

Fig 6. Ability to Evaluate Question Processing across Stages by Grade

	Comprehension	Retrieval	Judgment	Response	Overall
Grade 4	2.13	2.08	1.62*	2.27	2.58
Grade 8	2.10	2.15	1.83*	2.35	2.79
Total	2.11	2.11	1.72	2.31	2.69

Verbal Reports Completeness

- We examined whether there are differences in the completeness of verbal reports regarding inclusion of information illustrating four stages of the question response process (3-point Likert Scale: 1 – No, 2 – Somewhat, 3 – Yes).
- Results showed no differences between grade 4 and grade 8 students in their completeness of the verbal reports in terms of information relevant to comprehension, retrieval, judgment and response processes.

Analysis

Student verbal reports were coded by three coders. The following measures were included to evaluate verbal reports quality and completeness:

- Verbal report length: number of words;
- Duration of the think-aloud (seconds)
- Verbal report quality score based on the following criteria:
 - Relevance of verbal reports content to the question tested on a 3-point Likert Scale (1 – No, 2 – Somewhat, 3 – Yes).
 - Completeness of the verbal report content in terms of including information illustrating four stages of the question response on a 3-point Likert Scale (1 – No, 2 – Somewhat, 3 – Yes).
 - Ability to evaluate question processing at each of four response stages: comprehension, retrieval, judgment, and response on 3-point Likert Scale (1 – Unable, 2 – Somewhat, 3 – Able).

- Results shows that the think-aloud method does not vary in the quality of the verbal output obtained from children in the two age groups.
- The think aloud method is equally effective in producing verbal reports of sufficient quality and content to evaluate question processing and performance in both Grade 4 and Grade 8 students.
- Although one might expect to find that 4th graders are more likely to verbalize information that is not related to the topic of the question being tested, results show there is no observable difference in the relevance of the topic of the student think-aloud between the two grades.
- The verbal reports from both groups did not differ in terms of inclusion of information that pertains to stages of questions response processing: comprehension, retrieval, judgment, and response .
- However, it may be more difficult to identify and evaluate information related to the Judgement stage in verbal reports from 4th grade students compared to 8th grade, although this stage is, in general, more difficult to evaluate compared to other processing stages.

Conrad, F. G. and Blair, J. (2001, August 5-9). *Interpreting verbal reports in cognitive interviews: Probes matter*. Proceedings from the Annual Meeting of the American Statistical Association. Alexandria, VA: American Statistical Association.

de Leeuw, E., Borgers, N., and Smits, A. (2004). Pretesting questionnaires for children and adolescents. In Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E. Martin, J., and Singer, E. (Eds) *Methods for testing and evaluating survey questionnaires* (pp. 409 – 429). Hoboken, NJ: John Wiley and Sons, Inc.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

van Someren, M. W., Barnard, Y. F., and Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London, England: Academic Press.

More Information:

Please send questions or feedback to: Mila Sugovic, sugovicm@eurekafacts.com.