PEARSON

technical report

Cognitive Labs

Sasha Zucker

Christy Sassman

Betsy J. Case, Ph.D.

February 2004

Cognitive Labs

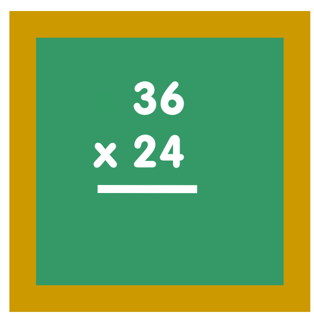
A cognitive lab is a method of studying the mental processes one uses when completing a task such as solving a mathematics problem or interpreting a passage of text. Developed and formalized using modern scientific research, cognitive labs have not only provided an effective insight into the functioning of the human mind, they also have been practically applied in the development of surveys, questionnaires, and assessments. Carefully executed cognitive labs yield valuable qualitative information that complements quantitative data from empirical research and statistical analysis. This qualitative data has repeatedly been demonstrated as invaluable in refining and validating assessments at both the early and late stages of development. As a leading publisher of high-quality assessment systems, Pearson Inc. (Pearson) has used cognitive labs as a regular activity in the development of educational assessments, including the Stanford Suite of Assessment Products and assessments for several state education agencies. This paper reviews the scientific background and practical applications of cognitive labs and examines their use at Pearson.

Verbal Reporting

The scientific basis for cognitive labs can be traced to recent developments in psychological research concerning the nature of mental processes. For much of the last century, modern psychological research in this area focused on collecting data that could be quantified, such as the directly observable behavior of a subject in a stimulus-response experiment. The directions of eye movements or the pushing of a button could be recorded and statistically analyzed in verifiable, objective ways (Ericsson and Simon, 1999, p. 1). This approach to research, *behaviorism*, supplanted *introspection*, the previously widely used method in which a researcher learns about the mental process of a subject by prompting the subject to talk about his or her own thoughts (Ericsson and Simon, p. 2).

The idea of someone revealing his or her own mental processes by talking about them—a practice known as “verbal reporting”—is familiar. People “think out loud” in everyday life, such as when a teacher asks a student to solve and explain a mathematics problem for the rest of the class or a leader is asked to justify the reasoning behind his or her decision (Ericsson and Simon, 1999, p. 78). Teachers use this strategy routinely to help students refine their problem-solving and computation skills. However, researchers were reluctant to accept verbal reporting as a scientific way of obtaining useful data about a person's mental processes. Verbal reports and the methods for obtaining them were considered unsystematic and therefore unreliable as a source of independently verifiable data, and the data itself was considered too subjective to interpretation by the researcher carrying out the experiment (Ericsson and Simon, p. 3). For example, while recording the subject's thoughts with a pen and paper, a researcher might omit "unimportant" data or present his or her own subjective interpretation of what the subject says as data (Ericsson and Simon, p. 4).

When verbal reports began to appear as data in contemporary scientific research, some scientists initially dismissed the data as a variant of introspection that is "useful for the discovery of psychological processes," but "worthless for verification" (Ericsson and Simon, 1999, p. 2). These doubts were overcome through subsequent research in which verbal reporting was refined and proven to be a formal and effective method for collecting data about mental processes. Researchers developed rigorous "thinking-aloud" protocols to formalize the observation and recording of data from verbal reporting as they would for any other observable behavior (Ericsson and Simon, p. 3). Moreover, the use of audio and video recording devices combined with strict protocols for transcribing the verbal reports enabled researchers to collect data in an objective format (Ericsson and Simon, p. 4). As a result of these studies, verbal reports have regained their significance as a source of data about mental processes that can be understood in the same way as other types of observable behavior (Ericsson and Simon, p. 40).



Within the context of cognitive labs, verbal reports are either concurrent or retrospective. In a concurrent verbal report, the subject verbalizes thoughts as they are happening while remaining focused on completing a task (Ericsson and Simon, 1999, p. 16). For example, in multiplying 24 by 36 while giving a concurrent verbal report, a subject might say out loud "36 times 24," "4 times 6," "24," "4," "carry the 2," "12," "14," "144," and so on, until arriving at the answer, 864 (Ericsson and Simon, p. xiii). In a retrospective verbal report, the subject verbalizes the thoughts that he or she had while performing a task immediately after completing it (Ericsson and Simon, p. 16). Therefore, in a retrospective verbal report of the mathematics problem just given, a subject would describe his or her thought process for solving the problem immediately after arriving at the solution. Retrospective verbal reports may also allow subjects to verbalize their evaluative mental processes (Ericsson and Simon, p. 264), such as their perception of the difficulty of a mathematics problem or whether or not a passage of text held their interest. Research concerning these types of verbal reports has shown that both types of verbal reports are effective in providing accurate data about mental processes; in some cases, retrospective verbal reports provide more complete data, as they are less likely to affect the subject's performance of a task than concurrent verbal reports (Ericsson and Simon, p. xvi). Depending on the project and age of the subjects, Pearson's assessment specialists, research scientists, and psychometricians use concurrent, retrospective, or a combination of both data collection techniques.

Cognitive Labs in Assessment Research

The research that contributed to formalizing verbal reporting as a valid method for collecting data has subsequently found practical applications, especially in the development of assessments intended to measure differences among individuals (Ericsson and Simon, 1999, p. 40). During the 1980s, assessment specialists applied the verbal reporting research of psychologists to develop *cognitive labs* (also called "cognitive interviews") as an assessment research method (Willis, 1999, p. 1). In a cognitive lab, a student completes test items and verbally reports his or her thoughts related to the item using a combination of "think-aloud" sessions (concurrent verbal reporting) and interviews with the researcher after each item is completed (retrospective verbal reporting) (National Center for Education Statistics [NCES], 2002, p. 1). To obtain a more thorough understanding of the subject's mental processes, verbal reports are often combined with other behavioral data observed during the cognitive lab, such as how a subject uses a pencil and paper to solve a mathematics item.

In contrast to the traditional statistical methods of analyzing the validity and reliability of an assessment quantitatively, researchers have found greatest value in the qualitative nature of the data they obtain from cognitive labs. Examples of qualitative data include problems that emerge repeatedly to form a clear trend and "discoveries" that occur only a few times but are clearly important to the quality of the assessment (Willis, 1999, p. 28). Some examples of problems that are typically revealed by cognitive labs include items that are unclear and subject to different interpretations, instructions that are overly complicated or difficult to read, mismatches between an item and provided responses, and issues of bias and sensitivity (Wilson and Peterson, 1999, p. 999).

Developers of recent high-profile educational assessments have incorporated cognitive labs into their research and documented the resulting benefits extensively. In a 1996–97 study (NCES, 2002) to analyze and improve a survey concerning the backgrounds of students participating in the National Assessment of Educational Progress (NAEP), researchers relied on student participation in cognitive labs to reveal "item problems that would not otherwise have been detected," and "identify sources of … confusion and misunderstanding" (NCES, p.1). Cognitive labs were cited as effective for "the verification of an expected … problem or the discovery of a problem that was unanticipated" (NCES, p. 1). Moreover, cognitive labs were recognized as useful for finding solutions to problems with the assessment. Researchers viewed the process of carrying out the cognitive labs as a source for important insights as well as raw data.

During the development of the Voluntary National Test (VNT) in 1998, cognitive labs were used to contribute to the design of assessment items. In an evaluation of the development of the VNT, it was noted that:

> The evaluation of draft items … in one-on-one think-aloud sessions with 4th and 8th graders during May and June 1998—called cognitive laboratories—was a significant and innovative item development activity. … The cognitive labs are a potentially valuable tool for test development, providing direct feedback to the developers about student understanding of items. (Wise, Hauser, Mitchell, and Feuer, 1999, p. 10)

Cognitive labs bring value to the development of assessments by providing data that in some cases could not be discovered through traditional statistical analysis and by verifying quantitative findings in others. Moreover, the process of carrying out cognitive labs provides immeasurable insights to assessment developers. Clearly, profound improvements to an assessment's quality can be made through the use of well-executed cognitive labs.

Use of Cognitive Labs at Pearson

Cognitive labs have played an important role in the development of Pearson's Stanford Suite of Products, especially the *Stanford Achievement Test* Series, Tenth Edition (Stanford 10). The data and insights gained from these cognitive labs, combined with other scientifically based research and development methods, such as focus groups and statistical analyses, prove to be invaluable for developing assessments to the highest standards of quality. In the early stages of an assessment's development, Pearson uses cognitive labs to gather data for the development of new items, forms, improved navigation layout, and prototype tests. In the last stages of an assessment's development, Pearson frequently relies on cognitive labs to determine the causes of statistical abnormalities in items and to further identify needed improvements to test design that cannot be indicated by statistical data alone. Pearson test developers agree that cognitive labs provide them with invaluable experiences that are crucial to their development of high-quality educational assessments.

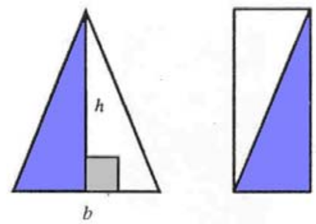
The design of each cognitive lab study used by Pearson is customized to the nature of the assessment and the informational needs of the developers. Typically, a cognitive lab consists of an interviewer administering items to a student in one-on-one sessions. Before research begins, a protocol for the cognitive lab is carefully designed to guide interviewers in eliciting and recording student responses accurately. In addition to following protocol, interviewers are trained to avoid influencing the student's response through verbal or visual cues, such as offering assistance or indicating whether the student's response was correct.

Students respond to each item while the interviewer records observations on the student's behavior in an interview booklet. A cognitive lab interview booklet is designed to guide the interviewer's record of observed behavioral data according to the cognitive lab's protocol and may include checklists of the possible behaviors that the student could exhibit while responding to an item (see Figure 1). To verify the objectivity of these written observations, a video or audio recording of each one-on-one session is always made for later review and verification by assessment specialists.

After the student completes each item, the interviewer may guide a retrospective verbal report from the student through a series of probing questions. To gather data about the thought process behind the student's response, the interview booklet may include questions to be asked of the student immediately after he or she completes the item ("How did you get your answer to this question? What made you do that?"). The interviewer's questions may also prompt evaluative verbal reports. Different sequences of questions are suggested for items answered incorrectly, left incomplete, or omitted. When the student has completed the assessment, the interviewer may also ask general questions about it. Retrospective verbal reports are typically used for the lower-grade-level students (K–3).

Because of the time required to complete a cognitive lab study, smaller numbers of students participate in a cognitive lab study than during an assessment's standardization and normalization studies. More than 1,300 students participated in cognitive labs around the nation in a Pearson study concerning the document design and layout of Stanford 10 (Case, 2003, p. 3). Several hundred students participated in cognitive labs on other Pearson products during 2003. Pearson recruits cognitive lab participants representative of the student population to whom the assessment will be administered.

10c) The rectangle below was built by cutting the triangle pictured into two pieces and then using them to form the rectangle. Write the formula for the area of the rectangle in terms of b and h. Provide an explanation that justifies your formula.



Intended Mathematics: Relate the area of a triangle to the area of a rectangle ______ Used ______ Not Used

Observation Period:

Additional Materials: ____Calculator (AM1) ____ Textbooks (AM2) ____Notebooks (AM3) ____Portfolio (AM4)

____Manipulative(s) ________________________(AM5) _____ Other ____________________________ (AM6)

Interview Period:

1) **Tell me in words what you did to get your answer to this problem.**
Probe as needed with: **What in the question made you do that?**

Based on the student's verbal explanation, mark as many of the following as apply:
_____ Looked at the graphics, diagram, table, or graph (SR1)
_____ Recalled prior knowledge (SR2)
_____ Used simple mathematical operation (SR3)
_____ Just "knew" (SR4)
_____ Used multiple steps (SR5)
_____ Obtained significant information from the graphic, table, or graph (SR6)
_____ Discussed the underlying mathematical concepts (SR7)
_____ Examined patterns and structures to detect regularities (SR8)
_____ Proved or disproved conjectures (SR9)
_____ Formulated generalization (SR10)
_____ Discussed the limitations of a generalization (SR11)
_____ Developed or explained arguments about mathematical concepts (SR12)
_____ Planned as part of solution process (SR13)
_____ Integrated a number of concepts and skills to solve the problem (SR14)
_____ Used an understanding developed in a previous question {indicates learning is taking place} (SR15)
_____ Relied on previous question {indicates item dependence} (SR16)
_____ Other (SR17) … Describe:

Figure 1. Sample interview booklet page from a Pearson cognitive lab for middle school students (size reduced for this report).

2) **Was this question easy, medium, or hard for you?** (Circle one) **Why do you say it was** (fill in student response)?

3) If a student inaccurately solved, did not completely solve, or skipped the problem, ask the following questions as they apply. Use the answers given and probe as needed to determine the cause of the problem and code appropriately.

- If item is obviously wrong, probe for source/reasoning behind the error.
- If item is incomplete, ask: **What did you do first?** and **Why did you stop?**
- If item was omitted, ask: **Why did you not do this question?**
  Probe to determine what the student did understand and at what point in his/her thinking he/she decided not to attempt to answer.

Scribing Area :

Figure 1. Sample interview booklet page from a Pearson cognitive lab for middle school students (continued - size reduced for this report).

Assessment specialists perform cognitive lab studies at the students' school when possible or at Pearson's facilities (for students in grades 3 and higher). Studies may also occur locally or nationally at the schools that participating students attend. In national studies, an interviewer trained in the cognitive study protocol may perform the cognitive lab with verification by an assessment specialist or psychologist using the video or audio recording.

The data collected during cognitive labs can be used to evaluate test items in both the early and the late stages of an assessment's design. In the early phases of the development of an assessment, before field testing, interviewers can clearly identify items that students repeatedly answer incorrectly, leave incomplete, or omit. Discovering such reliability issues well before they are revealed through costly large-scale field testing and time-consuming statistical analysis improves item design. Cognitive lab data also improve test validity through the early identification of items too difficult for the intended level of a form. Significant improvements are made to the layout of forms through the isolation of problematic text and graphics that lead students to respond incorrectly. Issues similar to those above are frequently revealed during cognitive labs.

In addition to revealing problems, cognitive labs can be used to inform decisions regarding the inclusion, exclusion, or revision of innovative, experimental item designs and approaches to assessment. For example, to investigate the impact of the use of full color in Stanford 10 test materials, assessment specialists used cognitive labs to ascertain student reactions and comments. The study revealed strong positive reactions from students who commented that the use of color made the test more interesting, captured their attention, and improved visibility of items compared to a black-and-white test. Color was frequently cited as the aspect of the test to which students had the strongest positive reaction. The incentives for using color in Stanford 10 were verified and strengthened by the cognitive lab results.

In the later stages of assessment development, after statistical analysis has identified undesirable characteristics in an item, cognitive labs are effective in determining the underlying cause of item issues as well as providing insights into the corresponding solutions. Often, interviewers can immediately identify a potential solution to an item's problem through the student's retrospective reports and observed behavior. For example, during the try-out stage of a Pearsont mathematics assessment, students consistently responded incorrectly to an item that included graphs. A cognitive lab study of the item revealed that the students were confusing the item's graph with a graph printed on the following page which was visible because of the test booklet paper's weight. Interpreting the wrong graph caused the students to respond incorrectly. The source of the irregularity was found and solved by increasing the paper's weight.

Data collected during cognitive lab studies have also been applied during the advanced stages of assessment development. In assessments constructed with open-ended items, the video and audio recordings from the cognitive lab can provide the actual responses used as the basis for scoring rubrics and also can serve as a reference for the standard-setting committees that determine an assessment's performance levels.

The benefits of using cognitive labs are well documented through their use in the research and development of Pearson products. Pearson will continue to rely on cognitive labs to enhance the validity, reliability, and overall quality of its educational assessment products.

Conclusion

Through their application in the field of educational measurement, cognitive labs provide clear benefits and offer a significant return on invested resources. The qualitative nature of the data from cognitive labs complements and enhances the statistical, quantitative data of traditional research methods used in the development of large-scale assessments. Pearson's use of cognitive labs during the creation of assessments such as Stanford 10 and clinical products demonstrates the incorporation of innovative, scientifically-based methods of research and development. By effectively using cognitive labs with other scientifically sound methods, Pearson establishes a commitment to quality that is second to none and will continue to lead assessment publishers in producing high-quality, valid, reliable assessments.

References

Case, B. J. (2003). *Test and answer document design and layout.* San Antonio, TX: Pearson Inc.

Ericsson, K. A., & Simon, H. A. (1999). *Protocol analysis: Verbal reports as data.* Cambridge, MA: Massachusetts Institute of Technology.

National Center for Education Statistics. (2002). *The measurement of instructional background indicators: Cognitive laboratory investigations of the responses of fourth and eighth grade students and teachers to questionnaire items.* Washington, DC: U.S. Department of Education.

Willis, G. B. (1999). *Cognitive interviewing: A "how to" guide.* Research Triangle Park, NC: Research Triangle Institute. Retrieved January 21, 2004, from http://appliedresearch.cancer.gov/areas/cognitive/interview.pdf

Wilson, B. F., & Peterson, L. S. (1999). *Using the NCHS cognitive lab to help design cycle VI of the national survey of family growth.* Proceedings of the Survey Research Methods Section, American Statistical Association (2000), 997–1002. Retrieved January 21, 2004, from http://www.amstat.org/sections/srms/Proceedings/papers/1999_174.pdf

Wise, L. L., Hauser, R. M., Mitchell, K. J., & Feuer, M. J. (1999). *Evaluation of the voluntary national tests*. Washington, DC: National Academy Press.

Additional copies of this and related documents are available from:
Pearson Inc.
19500 Bulverde Rd.
San Antonio, TX 78259
1-800-211-8378
1-877-576-1816 (fax)
http://www.pearsonassess.com