Published in final edited form as:
Soc Work Res. 2008 January 1; 32(1): 18.

Cognitive Testing and the Validity of Child-Report Data from the Elementary School Success Profile

Natasha K. Bowen
School of Social Work, University of North Carolina at Chapel Hill

Abstract

The Elementary School Success Profile (ESSP) is a social environmental assessment tool that collects data from 3rd, 4th, and 5th graders, their parents/guardians, and their teachers. Ensuring the validity of the data collected with the child report component was a primary concern during the development of the ESSP. This article describes how cognitive testing was used to promote the validity of ESSP child report data. Four types of response problems were identified in data collected from 58 children. Strategies to address problems are described, and evidence of the association between item modifications and improved performance is presented. Cognitive testing is recommended as a standard procedure in the development of child report instruments.

Keywords

Measurement; Validity; Elementary School; Social Environment; Cognitive Testing

The Elementary School Success Profile (ESSP, N. K. Bowen, Bowen, & Woolley, 2004) is an online ecological assessment tool designed to help school-based practitioners identify social environmental influences related to school success among 3rd through 5th graders. The ESSP provides practitioners with information on a range of social environmental factors that are known to influence developmental outcomes, including social behavior and academic performance. Instead of simply *describing* problematic behavior or performance like many school assessment tools, the ESSP provides insights into potential *causes* of those problems. ESSP results help practitioners identify appropriate social environmental intervention targets at both the individual and group levels.

The ESSP is also unique among elementary school assessments in terms of how it was developed. This article focuses on how cognitive testing with children was used to promote the validity of ESSP data collected from children. The article describes how cognitive testing was conducted, what problems it revealed, how those problems were addressed, and how cognitive testing increased the confidence practitioners can have in the validity of ESSP data. Recommendations on how to conduct cognitive testing are presented for researchers developing new instruments for children.

The ESSP

Three separate online components of the ESSP collect data from children, parents/guardians, and teachers, respectively. Numerous features of The ESSP for Children promote the reliability and validity of data collected: e.g., graphics and animations to maintain children's attention;

Natasha K. Bowen, Assistant Professor, School of Social Work, University of North Carolina, Chapel Hill, NC 27599-3550, Phone: (919) 843-2434, nbowen@email.unc.edu, Fax: (919) 843-8715.

screens explaining how to complete the ESSP; and an audio option to overcome reading difficulties. It takes most children about 15 minutes to complete the instrument's 83 items. The ESSP for Children is available in English and Spanish. The 12 dimensions that children report on are: Neighbors Who Care, Teachers Who Care, School is a Fun Place to Learn, School is a Fun Place to be with Other Children, Friends Who Care, Accepted by Peers, Friends Have Good Behavior, Family Who Care, Good Physical Health, Good Adjustment, Positive Feelings about Self, and Knows Where to Get Support.

The ESSP was developed in response to requests from school-based practitioners using the School Success Profile (SSP, G. L. Bowen, Richman, & Bowen, 2002; G. L. Bowen, Rose, & Bowen, 2005). The SSP was developed in the early 1990's by G. L. Bowen and J. M. Richman and has been used to assess over 50,000 middle and high school students. The SSP is a well-validated self-report instrument that asks adolescents about their perceptions of their neighborhood, school, peer system, and family, as well as their physical and psychological well-being and school performance. These domains are also assessed with the ESSP, but the number and nature of dimensions within the ESSP domains differ, wording of items within similar dimensions differ, and data are collected from parents and teachers as well as children.

Validity

Validity most commonly refers to "how well [an instrument] measures what it was designed to measure" (Corcoran & Fischer, 2000, p. 16) or, in the case of one scale, whether or not the "latent variable shared by items is, in fact, the variable of interest to the scale developer" (DeVellis, 2003, p. 49). Often validity is assessed indirectly in terms of how well items on a scale reflect the major facets of a construct (content validity), and how strongly scores obtained with scale items relate to scores from other measures of the same construct (criterion-related validity) or theoretically related constructs (construct validity). These measures of validity exclude an important direct measure of score validity--feedback from respondents about what their responses mean (N. K. Bowen, in press; Woolley, Bowen, & Bowen, 2004). Cognitive testing provides this type of direct assessment of data collected with instruments.

Cognitive Testing

Cognitive testing in scale development involves the use of questioning techniques to understand the thought processes of respondents while they respond to questionnaire items. It is used to "investigate the total question-answer process and discover sources of confusion and misunderstanding" (de Leeuw, Borgers, & Smits, 2004, p. 423). When used iteratively—i.e., when items are retested after being revised based on cognitive testing findings—it can also provide evidence that item performance has improved.

Because children's social, emotional, linguistic, and cognitive skills differ significantly from those of the adults who design questionnaires, cognitive testing is an especially critical step in the development of child self-report instruments (Woolley, et al., 2004). Cognitive research conducted by Levine, Huberman, Allen, and Dubois (2001), for example, revealed unexpected problems and high error rates in 4th grade children's responses to simple questions about their homes and parents. Their study also documented reductions in those error rates after identified problems were addressed.

The literature on cognitive testing with children is sparse (de Leeuw et al., 2004), and even the general literature on cognitive testing provides little guidance about how to analyze cognitive data and address problems that are identified. Willis (2005) has noted that "the nature of the steps that cognitive interviewers take between data collection and decision-making is largely uncharted territory" (p. 151). In presenting the case that cognitive testing results support the

validity of data collected with the ESSP for Children, this article describes analysis and item improvement strategies and methodological recommendations for future cognitive testing.

Cognitive Testing as Part of the Development of the ESSP

During the development of the ESSP, three rounds of cognitive interviews were conducted with 58 children. Rounds 1 and 2 took place in 2001; Round 3 took place in 2003. Procedures and testing materials were developed by the author based on the literature on cognitive methods (e.g., Jobe & Mingay, 1989; Willis, Royston, & Bercini, 1991), relevant child development factors, the purpose of the study, and constraints imposed by the school settings in which the testing took place.

Data Collection

Because the ESSP for Children is designed to be administered to children using computers at school, cognitive testing took place on school computers. Each child tested between 11 and 21 of the 80 to 90 items on the ESSP child questionnaire in each round. Data collectors advanced through the program as necessary to the assigned testing questions for each child. In Rounds 1 and 2, teachers were trained to conduct the cognitive testing. A doctoral research assistant collected data in Round 3.

An interview process with concurrent probes was used (Jobe & Mingay, 1989). Children were asked questions as they read and responded to each questionnaire item. Children were first asked to read each item aloud. If children had trouble reading a word, they were asked if they understood the word when the data collector read it aloud. Second, children were asked to state the item in their own words. Then they were asked to select a response. Finally, a probe was used to elicit an explanation of why a particular response was chosen. The goal of the question sequence was to determine whether existing items were read and interpreted by children as intended by the researchers—a direct assessment of data validity. If items were not interpreted as intended, the goal was to determine (a) how items could be changed so they would be interpreted as intended, or, (b) if the adult-defined constructs could be altered to be consistent with child definitions.

Recording the Data

In Round 1 teachers were trained to complete a recording sheet while interviewing children. Teachers indicated on the sheet if a child had “no difficulty,” “minor difficulty,” or “major difficulty” reading an item. If the child had difficulty with any word, the word was recorded. Teachers also indicated if the child understood the problem word when the teacher read it aloud. After asking the child to put the question into his or her words, the teacher indicated on the recording sheet if the child understood the question. If the child did not understand a question, the teacher summarized the child’s erroneous interpretation in writing. Finally, after the child was asked to select and explain his or her answer choice for the item, the teacher summarized the response. This documentation method minimized the recording burden of the teachers collecting the cognitive data, exploited the teachers’ existing skills in assessing reading and comprehension skills, and avoided a number of pitfalls of note-taking described by Willis (2005). Because of the richness of the findings from this round, however, audio-taping was used in subsequent rounds to obtain verbatim data. Audio-tapes were transcribed for analysis.

Data Analysis during the Development of the ESSP

The purpose of data analysis during the development of the ESSP was to promote validity of ESSP child data by addressing any threats to validity that were revealed. We were interested in what could be learned from all problems, even those that occurred for just one child. Cognitive testing data were examined systematically after each data collection by the author

and one to two other researchers. All data collected on each item were examined simultaneously for incidents and patterns of problems. Four categories of problems emerged inductively from the data: (a) word recognition problems, (b) comprehension problems, (c) incongruence between response choices and explanations of the choices, and (d) misapplication of response options to item content. How problems were categorized is described in more detail below.

The Current Analyses

The purpose of the current analyses was to examine data from all three rounds of cognitive testing simultaneously and retrospectively to test the hypothesis that cognitive testing promoted the validity of the ESSP for Children. The re-analysis was conducted by the author using recording sheets from Round 1, transcribed interview data from Rounds 1 and 2, and successive versions of the ESSP for Children documenting the history of items from Round 1 to Round 3.

An item was considered to have a word recognition problem when a child did not recognize a word or misread a word in the item without self-correction. Comprehension was considered to be a problem when a child misunderstood *who* or *what* a question was about, or the intent of the item.

Most items on the ESSP for Children use the response set “never,” “sometimes,” “often,” and “always.” Response congruence was considered to be a problem when a response choice (e.g., “sometimes”) was not logically consistent with the language used to explain the response choice (e.g., “because that happens all the time”). Response congruence was primarily related to the child’s understanding of the response options. Assessment of the appropriate *application* of response options to a particular item, in contrast, involved an evaluation of whether the child applied the response options to the proper concept in the item. For example, in response to the item, “I can talk to grownups at my school when I need help,” one child justified his response of “sometimes,” by saying, “sometimes I can get the answer by myself and sometimes I can’t.” The response indicates understanding of the frequency options, but misunderstanding of the concept of interest in the item. This type of problem could be classified as a comprehension problem, but because it relates to the simultaneous processing of response options and item content, we treat it as a separate category.

The data re-analysis resulted in descriptions of the number and nature of problems revealed with cognitive testing, classification of the strategies used to address problems, and documentation of the performance of revised items in subsequent rounds. The re-analysis also allowed us to document what ESSP constructs mean to children in their own words. Recommendations for researchers/practitioners interested in designing and testing child questionnaire items also emerged from the analysis.

Results of Cognitive Testing Data Re-Analysis

Between 16 and 23 children took part in each round of cognitive testing. Each sample was a convenience sample drawn from a different elementary school identified through staff in a private statewide after-school program. As shown in Table 1, African American children, 3rd graders, and children with below average reading levels are over-represented in the sample, especially in rounds 1 and 2, due to concern that the ESSP collect valid data from children from these groups.

Testing and Problem Identification Statistics

Table 2 summarizes testing and problem data for the three rounds of the cognitive testing. Cognitive testing revealed a substantial number of problems with items in each round of testing.

The testing procedures revealed numerous problems affecting the validity of ESSP data in each round.

Types of Problems, Strategies to Improve Items, and Benefits of Revisions

Table 3 presents descriptive data on the four types of problems identified with cognitive testing. Overall, comprehension problems were most common, and word recognition problem were least common.

Word Recognition Problems

Examples: As shown in Table 3, word recognition problems occurred in a small number of cases in each round. Table 4 presents the words that caused children difficulties. All of the problematic words were understood by the children when read aloud by the data collector.

Strategies: Three strategies for addressing the threat to validity posed by word recognition problems were identified: (a) deletion of the problem word, (b) addition of a definition to the help screen for all items containing the word, and (c) reliance on the audio feature. Strategies (b) and (c) were both used for any problem word that was retained. The fact that all problem words were understood by children when they were read out loud, suggests that the audio option in particular will mitigate any remaining word recognition threat to the validity of ESSP data.

Benefits: The performance in subsequent rounds of many items that were reworded to eliminate problem words could be examined. After the word “bother” was replaced with the expression “pick on” in three items about children being harassed, 0 of 29 testers of the three items had trouble reading the new expression. After the item “I have other aches or pains or feel sick” was simplified to “I feel sick,” 0 of 6 children had trouble reading the item. After Round 1, the word “confused” was removed from an item that read “Do you ever feel lost and confused?” No word recognition problems occurred among the 10 subsequent testers of the revised item, which read, “Do you ever feel like you don’t know what to do?” After “advice” and “judging” were removed from one item, 0 of 5 testers had difficulty recognizing words in the new item. When the expression “provide you with help” was replaced with “help you out” in another item, 0 of 5 testers had trouble with the item. The word “grownups” was replaced by “adults” in multiple items before Round 3. Out of 49 children who tested items with the word “adult” in them in Round 3, none had difficulty with the word.

Misunderstanding Content

Examples: As shown in Table 3, a substantial proportion of validity problems revealed with cognitive testing were comprehension problems. This finding is consistent with other cognitive testing studies reviewed by Willis (2005). Some comprehension problems resulted from misunderstanding of *who* a question was asking about. For example, in response to the item, “When I need help, I can talk to adults who live near me,” one child referred to the availability of his mother (not neighbors). Another, in response to the item, “I have friends to talk to and play with on weekends,” responded that her *brother* “is always picking on me and he doesn’t like to play with me that much.”

In other cases, children did not understand the main intent of items. For example, in response to an item evaluating the availability of friends to play with “outside of school” (i.e., when children were not at school), one child referred to having friends to play with outside *at* school during recess. In response to the item, “I would rather play by myself than with friends,” two children failed to respond to the preference concept. One responded that when friends were not available, he could play by himself. Another simply said she did not play by herself. In another example, a child answered “sometimes” to the item, “My friends hit other kids,” an item designed to assess peer antisocial behavior. The child’s explanation—“[because]

sometimes when playing they hit each other"—suggested that he was not reporting antisocial behavior. In response to the item, "I feel I can do things as well as most other kids," another child responded, "Sometimes I don't want to do what others do." Two children responded to the self esteem item, "I feel happy with myself," in terms of how often they were in a happy mood. Other children failed to respond to *any* relevant concept in the items. For example, in response to one social support scale item assessing whether children perceived someone being "on their side," one child said "sometimes kids they be getting in arguments, they start fighting, and other kids jump into it …."

Strategies: Four strategies were used to address the threat to validity posed by comprehension problems. The first strategy--use of text to introduce an upcoming question--has been found in previous research to increase the quality of children's responses to items (Borgers and Hox, 2002, as cited in de Leeuw et al., 2004). To address one child's reference to his brother in the item asking about friends, a screen was added before the sequence of friend questions in the online questionnaire stating: "Friends are the kids you talk to and play with. Don't count brothers and sisters." A second strategy was deleting problematic items.

A third strategy was reordering scale items to clarify item meaning. This strategy was used in response to the child who reported on playful hitting instead of antisocial hitting. The items in a scale assessing friends' behavior were reordered after Round 3, so that all negative items were grouped separately from the positively worded items. Having the hitting item embedded among items assessing fighting, being mean, and lying is expected to prevent confusion about what kind of hitting to report on.

The fourth strategy—simplifying items—was used to address many comprehension problems. As discussed by Woolley et al. (2004), the ability of a child to respond validly to a self-report item is increased when the cognitive demands of the items are minimized. Therefore, any characteristic of an item that increases the cognitive demands of the item, such as advanced vocabulary, greater length, conditional statements, time references, and a higher level of abstraction in concepts, increases the likelihood that a child's response will not be valid. In an example of how eight Knows Where to Get Support items were simplified, one item changed from "During the last week, how often did someone tell you that they are on your side, even if they may not agree with you?, to "How often does someone tell you they are on your side?" and then to "Someone tells me they are on my side."

Simplification was also used with a set of Positive Feelings about Self questions. In Round 1, global self esteem items from the SSP for adolescents, such as, "I feel I can do some things very well," and "I feel I can do things as well as most other kids" caused numerous comprehension problems. Simplification involved making the items more concrete and domain specific, for example, "I am good at art," "I am good at sports".

Benefits: The performance in subsequent rounds of items that had simplified to address comprehension problems could be examined. A representative selection of examples is provided. After Round 1, the item, "I have friends to talk to and play with outside of school" was made into two new items about the availability of friends to play with after school and on weekends. Out of 18 tests in Rounds 2 and 3, only one comprehension problem occurred, and it was related to considering a sibling a friend, not the setting in which friends were found.

In Round 2, six children experienced 13 comprehension problems across 47 tests of the eight cognitively demanding Knows Where to Get Support items (in addition to eight problems of other types). In Round 3, after the items were greatly simplified, there was only one comprehension problem across 35 tests by five children. Simplification of the Positive Feelings about Self scale was also associated with its improved performance. In Round 1, four of four

testers demonstrated a total of nine comprehension problems across six items. In Round 3, after all items had been revised, 0 of 5 testers of the self esteem scale items demonstrated comprehension problems. About half of the explanations given for responses to the new self esteem items, however, related to how often children engaged in an activity, how much they liked the activity, or how hard they worked at it. This example may reflect a point made by Willis (2005): "there may be no manner in which we can feasibly modify questions such that they satisfy our measurement objectives" (p. 155). Promoting consistency between intended constructs and respondents' meanings may at times involve refining construct definitions to match, in this case, developmental reality.

Response Option Incongruence

Examples: As shown in Table 3, response option incongruence was a problem of different proportions in each round of cognitive testing. One child indicated that grownups in the neighborhood were nice to him "often." In explaining his response, however, he said, "because sometimes they do and sometimes they don't." Another child said "sometimes" he could talk to neighbors when he needed help, but explained his answer by saying "most of the time I can talk to them." Another child said he and his friends "often" do fun things together "because we always play fun games." Identifying incongruence was not always straightforward because a child could explain an "often" response in terms of why it was not an "always" response, or in terms of why it was not a "sometimes" response.

Strategies: In addition to cognitive data, our decisions about response options were guided by statistical issues (e.g., more response options are desirable, DeVellis, 2003) and developmental concerns (child items should have only 3 to 5 response options, de Leeuw, 2004; having multiple response sets could be confusing to children). The ESSP's 4-point ordinal frequency response option set was consistent with these guidelines and worked well in most cases. Therefore, to address the response option problems revealed in cognitive testing, two indirect solutions were used. First, a screen defining the four primary response options in the ESSP for Children was added to the computer program's introductory screen sequence. Second, the help screen for every item on the ESSP repeats the response option definitions.

Benefits: Our cognitive testing procedures focused only on individual items and not the introductory or help screens, so it was not possible to assess the benefits of these changes across the three rounds of testing.

Misapplying Response Options to Content

Examples: Another category of problem identified with cognitive testing was the misapplication of response options to item content. One child, for example, chose the option "never" for the item "grownups in the neighborhood would say something to me if I did something wrong." The child's response explanation indicated that "never" applied to how often he did something wrong, not how often neighbors would intervene. This category of problem differed from the previous category in that the response frequency chosen was appropriate to the explanation of how often a phenomenon happened, but it was applied to the wrong phenomenon.

Strategies: For the reasons stated above, changing response options was not considered as a strategy to address the misapplication of response options. Instead, strategies focused on simplifying content to reduce the cognitive demands of the items. Four social support items on which children had misapplied response options were simplified by removing a phrase specifying a time frame ("during the last week"), putting them in the present tense, eliminating word recognition problems, and shortening in other ways when possible. Woolley et al. (2004), hypothesized that the placement of conditional phrases, such as "when I need help,"

or “if I did something wrong” at the beginning of items might make the items less cognitively demanding than placing them at the end. Therefore, conditional expressions were moved to the beginning of two items from the Neighbors Who Care dimension for which children misapplied response options.

Benefits: In 20 tests of the four social support items that were simplified, no response application (or comprehension problems) occurred in Round 3. The two items in which conditional expressions were moved from the end to the beginning did not have response application problems in subsequent rounds, however, each had one comprehension problem. Further analysis indicated that across all three rounds, 10 items with conditional phrases at the beginning performed no better than 33 with the qualifying phrase at the end. Also across all rounds, items with conditional expressions were twice as likely to have problems (of all types) than items without them. It appears that conditional expressions make items more difficult regardless of where they are placed.

Discussion

Cognitive testing with children was used during the development of the ESSP for Children to promote the validity of data collected with instrument. A retrospective analysis of data from all three rounds was conducted to determine if cognitive testing accomplished this goal. The findings can increase school-based practitioners’ confidence in the quality of data collected with the ESSP for Children. The substantial number of problems found among items that had already undergone the careful scrutiny of numerous academic and education experts suggests the importance of cognitive testing as a routine method in the development of self-report items in quantitative instruments for children. As evidenced in the current study and in the Levine et al. (2001) presentation of cognitive testing data, in the absence of cognitive testing, even the simplest of items cannot be assumed to yield valid responses from children.

Four categories of problems were revealed in the cognitive data: word recognition problems, comprehension problems, response option incongruence, and misapplication of response options. Strategies to address problems included the deletion of problem words or items, simplification of items, moving conditional phrases, reordering scale items, and adding content to the online ESSP for Children’s support screens. Revised items that could be evaluated in subsequent rounds rarely had problems and only once had the same problem that had been addressed in the revision process. The iterative process of cognitive testing led to greater confidence that items in the final version of the ESSP for Children can be read, understood, and responded to in a valid way by 3rd through 5th graders, African American and White children with a range of reading abilities.

Cognitive testing also allowed us to document what each of the 12 ESSP for Children constructs mean to children, that is, to summarize the specific meaning of latent variables, based on the explanations of children. The findings are presented in Table 5. Practitioners using the ESSP receive a description of the meanings children give to scales on the ESSP for Children, which they can use as an aide to interpreting their results and planning interventions.

While a full discussion of findings and methodological aspects of cognitive testing of the ESSP for Children is beyond the scope of this article, the systematic re-analysis of cognitive testing data yielded guidelines for researcher and practitioners interested in developing or evaluating child report scales. First, the study provides substantial evidence that questionnaire items for children in middle childhood should be as short, simple in terms of vocabulary and structure, and concrete as possible. Time references and conditional phrases should be avoided. It is not possible to avoid every type of cognitive demand in child report instruments (e.g., a conditional phrase, or a vague word, such as “something”). School practitioners, however, should be aware

of the above threats to the validity of child-report data. While items containing any one cognitively demanding component may perform adequately for many upper elementary school students, the accumulation of demanding characteristics is likely to lead to invalid responses (see also, Woolley et al., 2004). Practitioners concerned about threats to the validity of data collected with their current tools might test items with a small number of children as described below, and/or take steps to address threats, such as reading items aloud to children, confirming comprehension, and allowing children to ask questions while completing child report instruments.

Second, for practitioners and researchers interested in employing cognitive testing in the development of their own child report instruments, the review of cognitive testing data yielded the following recommendations related to data collection. (More detailed recommendations are available from the author.)

Testing Situation

1. Test items in a setting and format as similar as possible to the intended administration conditions. This includes having children read or hear instructions and introductory text for items they are testing;
2. Whether they are research team members or school staff, data collectors should receive detailed training about what information is being sought in the question series and how to use follow-up probes to obtain that information.

Interview Questions

1. Ask children to read an item aloud. If they misread or cannot read a word, ask if they understand the word when they hear it said out loud;
2. Ask children to choose a response option for the item;
3. Ask children to explain their response choice;
4. Follow as necessary with probes to ensure that response choice explanations provide adequate information for assessing validity.

These recommendations omit the paraphrasing question used in ESSP procedures. The review of all rounds of ESSP cognitive testing revealed that regardless of how the paraphrasing question was worded, it yielded little useful data. De Leeuw (2004) has also made this observation. The best validity data were obtained with the question asking children to explain their response choices; this question should be the primary focus of the cognitive interview.

Valuable data were collected both by trained teachers and a research team member, and both with recording sheets and audio-taping. While fewer verbatim data are collected with recording sheets, ample valuable data can be obtained with well-designed forms. We transcribed audio-taped data, but Willis (2005) suggests that effective analysis can occur without that time-consuming step. Readers interested in applying an intensive, multi-rater analysis method for transcribed cognitive data are referred to Woolley, Bowen, and Bowen (2006). If the right interview questions are asked, therefore, it appears that researchers and practitioners can adapt their data collection and recording methods to local needs and constraints. In most applications of cognitive testing, however, multiple rounds will likely be needed to adequately assess and revise items. The literature does not provide guidance about when cognitive testing should end. While it is unlikely that all error can be avoided in item interpretation and response, it is not clear how much error is acceptable or unavoidable.

Limitations

Although extensive cognitive testing was conducted in the development of the ESSP for Children, additional rounds would have been beneficial to test item revisions made after Round 3. Although most revised items performed better in subsequent rounds, new problems continued to be identified in Round 3. Additional rounds would also have permitted an evaluation of the impact of strategies such as the audio feature and changes to introductory screens and help screens. Another limitation is the fact that multiple item improvement strategies were applied to some items before they were retested. No attempt was made to identify the impact of any one type of revision on future item performance. The aim of the re-analysis was simply to determine that previous problems did not recur.

As described by Willis (2005), small, non-random samples and convenience sampling preclude the statistical analysis of most cognitive data and limit the generalizability of findings. These factors are limitations of the current study. Our samples also differed across rounds in terms of reading ability, grade level, and race/ethnicity. Our overall sample size of 58, however, was reasonable for a qualitative study. Two major racial/ethnic groups were represented in the study's sample and 3rd graders and below average readers were over-represented. These features are study strengths, given the purpose of cognitive testing in the ESSP development process. A separate analysis, available from the author, suggested that improved performance was not explained by sample differences over time. Because children of all grade levels and reading abilities demonstrated problems with word recognition, content, or response options, it is recommended that future researchers conduct cognitive testing with the full range of ages and abilities targeted with the instruments being developed. Schools with diverse populations should be the source of the samples, and at least 15 to 20 children should be included in each round. To promote the equivalence of samples studied across rounds, a random sample of students representing the instrument's target population should be obtained and then randomly assigned to three or four groups for the successive rounds of testing.

Finally, the current study was not designed to determine if cognitive testing of ESSP items improved their *statistical* validity or reliability. As reported elsewhere (N. K. Bowen, 2006), the ESSP for Children has strong psychometric properties. It seems likely that eliminating unrecognized words, addressing sources of comprehension problems, and addressing response option problems contributed to the excellent statistical performance of the ESSP for Children, but the current study cannot prove that causal relationship.

Acknowledgments

The ESSP was developed in collaboration with Flying Bridge Technologies, with funding from the National Institutes of Health (NIH) and the National Institute on Drug Abuse (NIDA), grant numbers 1 R42 DA13865-01, 3 R41 DA13865-01S1, and 2 R42DA013865-02. The second two rounds of cognitive testing were supported by a Junior Faculty Development Award to the author from the University of North Carolina at Chapel Hill. Findings, opinions, and recommendations expressed in this article are those of the authors and not necessarily those of Flying Bridge Technologies, NIH, or NIDA.

References

Bowen, GL.; Richman, JM.; Bowen, NK. The School Success Profile: A results management approach to assessment and intervention planning. In: Roberts, AR.; Greene, GJ., editors. Social Workers' Desk Reference. New York: Oxford University Press; 2002. p. 787-793.

Bowen, GL.; Rose, RA.; Bowen, NK. Reliability and validity of the School Success Profile. Philadelphia, PA: Xlibris; 2005.

Bowen NK. Psychometric properties of the Elementary School Success Profile for Children. Social Work Research 2006;30(1):51–63.

Bowen NK. Validation. International Encyclopedia of the Social Sciences. in press.

Bowen, NK.; Bowen, GL.; Woolley, ME. Constructing and validating assessment tools for school-based practitioners: The Elementary School Success Profile. In: Roberts, AR.; Yeager, KR., editors. Evidence-based practice manual: Research and outcome measures in health and human services. New York: Oxford University Press; 2004. p. 509-517.

Corcoran, K.; Fischer, J. Measures for clinical practice: A sourcebook. 3. Vol. 1. New York: Free Press; 2000.

de Leeuw, E.; Borgers, N.; Smits, A. Pretesting questionnaires for children and adolescents. In: Presser, S.; Rothgeb, JM.; Couper, MP.; Lessler, JT.; Martin, E.; Martin, J.; Singer, E., editors. Methods for testing and evaluating survey questionnaires. Hoboken, N.J: John Wiley & Sons; 2004. p. 409-429.

DeVellis, RF. Scale development: Theory and applications. 2. Thousand Oaks, CA: Sage; 2003.

Jobe JB, Mingay DJ. Cognitive research improves questionnaires. American Journal of Public Health 1989;79(8):1053–1055. [PubMed: 2751028]

Levine, R.; Huberman, M.; Allen, J.; DuBois, P. The measurement of home background indicators: Cognitive laboratory investigations of the responses of fourth and eighth graders to questionnaire items and parental assessment of the invasiveness of these items, NCES 2001–19. Washington D.C: U.S. Department of Education, National Center for Education Statistics; 2001.

Willis, GB. Cognitive interviewing: A tool for improving questionnaire design. Thousand Oaks, CA: Sage Publications; 2005.

Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. Applied Cognitive Psychology 1991;5:251–267.

Woolley ME, Bowen GL, Bowen NK. Cognitive pretesting and the developmental validity of child self-report instruments: Theory and applications. Research on Social Work Practice 2004;14(3):191–200.

Woolley ME, Bowen GL, Bowen NK. The development and evaluation of procedures to assess child self-report item validity. Journal of Educational and Psychological Measurement 2006;66:687–700.

Table 1

Characteristics of the Sample for Each Round of Testing and the Total Sample: Percentage (*n*)

	Round 1 (*n*=16)	**Round 2 (*n*=23)**	**Round 3 (*n*=19)**	**Total (*n*=58)**
Reading Level				
Below	68.8 (11)	17.4 (4)	15.8 (3)	31.0 (18)
Average	31.3 (5)	47.8 (11)	73.7 (14)	51.7 (30)
Above	0.0 (0)	34.8 (8)	10.5 (2)	17.2 (10)
Race/ethnicity				
African American	100.0 (16)	60.9 (14)	5.3 (1)	53.4 (31)
White	0.0 (0)	39.1 (9)	78.9 (15)	41.4 (24)
Hispanic/Latino	0.0 (0)	0.0 (0)	5.3 (1)	1.7 (1)
Mixed race	0.0 (0)	0.0 (0)	10.5 (2)	3.4 (2)
Grade Level				
3rd	100.0 (16)	56.5 (13)	42.1 (8)	63.8 (37)
4th	0.0 (0)	0.0 (0)	21.1 (4)	6.9 (4)
5th	0.0 (0)	43.5 (10)	36.8 (7)	29.3 (17)

Table 2

Summary of Test Characteristics and Problems Encountered in 3 Rounds of Cognitive Testing

	Round 1	Round 2	Round 3
Number of ESSP items tested	89	93	81
Number of child testers	16	23	19
Number of item tests	228 (11 – 15 per tester)	339 (14 – 16 per tester)	382 (19 to 21 per tester)
Number of problems revealed	47	101	26
% of items causing problems	32.6 (29/89)	59.1 (55/93)	28.4 (23/81)
% of child testers demonstrating at least one problem	87.5 (14/16)	91.3 (21/23)	84.2 (16/19)
% of problems per item tests	21.9 (47/228)	29.8 (101/339)	6.8 (26/382)

Table 3

Prevalence of Four Types of Problems Revealed in Cognitive Testing

	Percentage of Total Problems		
	Round 1	**Round 2**	**Round 3**
Word recognition problems	12.8 (6/47)	10.9 (11/101)	3.8 (1/26)
Misunderstanding content	70.2 (33/47)	41.6 (42/101)	38.5 (10/26)
Response option incongruence	4.3 (2/47)	15.8 (16/101)	42.3 (11/26)
Misapplying response options to content	12.8 (6/47)	31.7 (32/101)	15.4 (4/26)

Table 4

Words That One or More Child Did Not Recognize in Each Round of Testing

Round 1: Aches, bother, confused, grownup, something
Round 2: Advice, encourage, headache, judging, need, provide, something, study, tired, toothache, worried
Round 3: Lending

Table 5

What ESSP Constructs Mean to Children According to Cognitive Testing Data

Domain & Dimension		Meaning of Construct
NGHHD	Neighbors Who Care	Neighbors are **friendly** (smile, wave, honk; let child visit, are friends of the child), **provide tangibles** (money for fundraisers, food, help with schoolwork); and **care** (available to help, can be trusted, are "there for me," take the child's side, watch over children to ensure safety)
SCHOOL	Teachers Who Care	Teacher is **friendly** (teacher and child get along, teacher is nice, teacher plays with child outside); **responsive** (even given that other children need to be called on and listened to, teacher helps, calls on, and listens to the child); **cares about schoolwork** (rewards good work and effort, helps child do better); Children also note that teacher "caring" is tied to their good behavior.
	A Fun Place to Learn	**Specific activities/subjects** are enjoyed, **learning** in general is fun and challenging, **teacher promotes fun and success** of students
	A Fun Place to be with Other Children	There are children who are **nice** and **like the child**, there are children to **play with,** and children the child **looks forward to seeing** at school
FRIENDS	Friends Who Care	Friends **respond** when child is upset or sad, are **happy for child** when he/she is happy, and engage in **fun activities** with child
	Accepted by Other Children	Reverse coded construct. Other children **say or do mean things**, **are jealous** of the child, **make fun** of the child, **exclude the child from play**.
	Friends Have Good Behavior	Friends **avoid antisocial behaviors** (hitting, fighting, hurting other children's feelings, telling lies); Friends **engage in prosocial behaviors** (help when someone is hurt, listen to teacher, help when asked by adults, do the "right" things, are kind). Children distinguish appropriately between negative behaviors and "getting mad" at friends, and between real and playful hitting.
FAMILY	Family Who Care	Caregivers or family show **unconditional caring** (give toys, take care of child when sick, say child is good, love child), give **conditional praise** (for schoolwork or helping behavior), **help child** (with chores or problems), **listen to child** (in general or about friends or problems), **engage in fun activities** with child. Children's responses reflect that caregivers are sometimes busy with work, other children, their own problems.
HEAL TH & WELL-BEING	Good Physical Health	Reverse coded construct. Children report on **somatic complaints** (stomachache, headache, feeling tired or sick), **dental**, **vision** and **hearing** problems. Sometimes children identify avoidable causes of problems (headache from ice cream, stomachache from eating too much, trouble sleeping because of scary movie).
	Positive Feelings about Self	Positive feelings about self are most often based on **specific skills** or **accomplishments** or **positive perceptions of activities**. Smart means they do well in school; good at art means doing well at certain art activities (but rarely all) or getting a good grade in art class; being proud of self is tied to performance in specific areas or to effort; good in music and sports means liking the activities, engaging often in the activities, or showing skill in the activities. Being nice or a good friend includes not being mean to or getting mad at friends. Children's responses indicate that they see both their strengths and weaknesses in all performance areas, and accept their weaknesses as part of life.
	Good Adjustment	Reverse coded construct. Good adjustment includes seeing **home as a safe place** where people care, listen, and help you. **Specific worries** include school performance, parents fighting, family members getting hurt, terrorism after 9/11. Not knowing what to do is related most often to **homework and schoolwork**. Responses indicate a perception that caregivers don't care about child when they are "mad" at child for misbehavior.

Domain & Dimension	Meaning of Construct
Knows Where to Get Support	There are **parents, friends, teachers, other school staff** who provide **emotional support** (listen, encourage children to do their best, ask how children feel, know what child is going through), and **tangible support** (homework help, lunch money).