

**Including English Language Learners in the Process of Test Development: A Study on
Instrument Linguistic Adaptation for Cognitive Validity**

Rachel Prosser

University of Colorado at Boulder

&

Guillermo Solano-Flores

University of Colorado at Boulder

Paper presented at the Annual Conference of the National Council of Measurement in Education,
Denver, Colorado, April 29-May 3, 2010

Abstract

This paper reports preliminary results from an investigation, still in progress, on the use of verbal protocols among native Spanish-speaking, English language learners (ELLs) of various proficiency levels and background characteristics. We focus on language use among ELLs during various stages of a cognitive interview designed to probe whether and how students benefitted from the inclusion of illustrations as a form of testing accommodation. While the majority of students did not use their native language, 29% of participants drew from their native language to convey their thoughts. These students varied considerably in their patterns of use of the two languages at different parts of the cognitive interviews. Our findings are consistent with research in the field of bilingualism. First, bilingual individuals vary tremendously in their patterns of use of two languages across different contexts. Second, bilingual individuals continually use their two languages when performing cognitive tasks, even if the tasks are given in only one of the languages and the individuals are expected to provide their responses only in that language. In addition, even ELLs who are classified as non- or limited-English proficient are capable of providing valuable information *in English* on their interpretation of test items. We discuss how these findings can be used to ensure the participation of ELLs in talk-aloud protocols as part of the cognitive validity procedures used in large-scale test development.

The purpose of this study is to document issues relevant to using verbal protocols with English language learners (ELLs) and to ensure proper inclusion of these students in cognitive validity studies during the process of test development. While the use of verbal protocols in testing is not new (e.g, Baxter, Elder, and Glaser, 1996; Baxter and Glaser, 1998; Hamilton, Nussbaum, and Snow, 1997), research on this important resource for test validation has focused on monolingual, native English speaking populations.

We address three important concerns in ELL testing. First, because cognitive research in education has been limited to mainstream populations, the extent to which findings from that research can be generalized to other groups remains uncertain (see Pellegrino, Chudowsky, & Glaser, 2001). Second, due to different communication styles, cultural and linguistic groups may vary considerably on the kind and amount of information they regard as necessary to give during verbal communication (Heath, 1993)—which may affect the interpretability of student verbalizations (Prosser & Solano-Flores, 2009). Third, ELLs are typically not included in the development stages of testing, which can potentially constitute a source of bias against them (Solano-Flores & Li, 2009).

In this paper, we report preliminary results from an investigation, still in progress, that examines how verbal protocols should be adapted for use with ELLs. We identify challenges that stem from the fact that ELLs may feel comfortable using both their first language (L1) and their second language (L2) to participate in cognitive interviews. Also, we investigate the minimum set of linguistic skills that verbal protocol administrators should have to effectively communicate with ELLs and obtain useful information using these instruments.

This investigation is being conducted as part of a larger National Science Foundation research project which examines the use of vignette-illustrations as a form of testing

accommodations for ELLs. As part of the activities performed for this larger research project, verbal protocols were used to probe whether and how native Spanish-speaking ELLs benefit from the inclusion of illustrations accompanying science test items with the intent to reduce limited proficiency in English as a source of measurement error.

We utilize verbal protocols with the intent to gain insight into the cognitive processes carried out by students taking items with and without illustrations. Intended to elicit thinking processes orally encoded by the participant (Ericsson and Simon, 1984, p. 63), verbal protocols have been thoroughly investigated as tools for probing understanding. Unfortunately, in spite of their value as tools for examining the cognitive validity of tests (Baxter, Elder, & Glaser, 1996; Ruiz-Primo, Shavelson, Li, & Schultz, 2001), their use with ELLs is yet to be investigated. While some research has examined the features of items that affect the validity of tests in investigations in which cognitive interviews are used with linguistically diverse students (e.g., Winter, Kopriva, Chen, & Emick, 2006), that research has not examined language use among ELLs during the cognitive interviews¹.

Previous research in the field of bilingual development has investigated important qualitative differences in the cognitive processes of bilinguals and monolinguals (Bialystok, 2002; Grosjean, 1989). Thus, we address the notion that the language system of bilingual individuals uses both the L1 and L2 for all cognitive processing. According to this model, during problem solving, ELLs may express their thoughts and build their arguments more effectively if they are allowed to use all of their linguistic resources—both L1 and L2—than if they are restricted to use only one language (Moschkovich, 2006).

Additionally, we address the tremendous heterogeneity of the bilingual populations. Grosjean (1989) describes bilinguals' language use by providing a continuum of the contexts in which

bilingual individuals function. At one end of the continuum are completely monolingual contexts in which bilinguals only use one language to communicate (the monolingual end). At the other end of the continuum are completely bilingual contexts in which both languages are continually mixed (bilingual end). Depending on context, a bilingual may find him/herself at any point on this continuum. In addition, due to individual factors, some bilinguals may prefer and promote completely monolingual contexts and rarely mix languages in their utterances. Other bilinguals may prefer and promote bilingual contexts, preferring to code-switch during most interactions.

Our procedures are based on the premise from the field of sociolinguistics that bilingual populations are not homogenous, and that the use of the L1 or L2 is shaped, among many other factors, by the topic of conversation (Fishman, 1965). In testing, this notion is reflected by the fact that ELLs' performance is inconsistent across tasks and languages (Solano-Flores, 2006; Solano-Flores & Li, 2009); every ELL student has a unique set of strengths and weaknesses in L1 and in L2 and each task poses a unique set of linguistic demands. Some students may be more comfortable speaking about one topic in L1 and another topic in L2; others may feel comfortable switching continuously between L1 and L2.

Recognizing this heterogeneity is important because, while some researchers (e.g., Kopriva, 2008) have suggested that it is important to collect background variables to identify trends among ELLs, others (e.g., Valdes and Figueroa, 1994), warn that broad background characteristics do not allow us to predict behavior at the microlevel of an individual's language use. That is, knowledge of facts such as the number of years students have lived in the U.S. or their age of acquisition (if that age can be determined at all) do not allow researchers to assume how students will use language at a given time during the cognitive interview. Therefore, we investigated language (English or Spanish) use among native Spanish speaking ELLs

participating in cognitive interviews. Also, we examined whether patterns in the use of one of the two languages or the use of both languages is consistent with variables commonly used to characterize the proficiency levels of ELLs—or whether these variables may even mislead researchers in their assumptions about how ELLs will use language in cognitive interviews.

Methods

Participants

In this study, 31 native Spanish-speaking ELLs participated from a middle school in a large school district in a western mountain state. Ten participants were in 6th grade, thirteen were in 7th grade, and eight were in 8th grade. Their ages ranged from 11 – 14 years. Eighteen participants were male and thirteen were female. Of the 31 students, four were classified as Non English Proficient (NEP), twenty six were classified as Limited English Proficient (LEP) and one was classified as Fully English Proficient (FEP)². Twelve students reported being born in the U.S., seventeen reported being born in Mexico, one student reported that he was born in Puerto Rico and one other reported being born in El Salvador. Students who were not born in the U.S. reported to have lived in the country from 1 – 13 years.

Protocol procedures

For the purposes of this investigation, the research team developed a verbal protocol to probe student reasoning when taking multiple choice items in science. The research team was most interested in how student reasoning differed for illustrated and non-illustrated items. For over a year, several iterations of a verbal protocol were developed in both English and Spanish to guide the researchers and maintain consistency across all stages of the interview: the introduction, warm-up items, concurrent reporting, retrospective reporting, and follow-up questions. The

development and piloting phases included only monolingual mainstream students due to limitations in the access to a large number of ELLs.

Modifications made to the protocol were informed by members of the technical advisory board, teachers who participated in the broader study, and piloting experiences. These modifications included adding information to the introduction to explain that researchers were interested in students' reasoning and that student participation was not for grade. Also, the researcher explicitly told students not to consider the researcher as a teacher or authoritative figure. In addition, members of the advisory board suggested that all follow-up questions be conducted after concurrent and retrospective reporting for all items took place as not to influence student reasoning. The Spanish version of the protocol was developed first by a native English-speaking bilingual researcher and later reviewed and modified by a native Spanish-speaking researcher to ensure it was in a Mexican dialect of Spanish, which we assumed to be the Spanish dialect most familiar to the majority of the ELLs participating in the study.

The interviewer visited the classroom of participating students for several weeks leading up to the interview. These initial meetings served as ice-breakers and allowed the researcher and students to become familiar with each other. Whenever possible, the researcher engaged in casual conversations with students, using both English and Spanish to provide contextual clues that the researcher was bilingual. In February 2010, the first author administered verbal protocols to the 31 ELLs. This researcher is a native English speaker who is fluent in Spanish.

Items

A total of sixteen 8th grade multiple choice items were used. These items were selected from various assessment programs, including CSAP (Colorado State Assessment Program), TIMSS (Trends in International Mathematics and Science Study), CST (California Standards Test), and

AIMS (Arizona Instrument to Measure Standards). In their original version, these items did not contain illustrations. An illustrated version of each of these sixteen items was created by adding an image next to the text of the item with the intent to provide non-linguistic support for ELLs. These images were designed according to the procedure described elsewhere (Solano-Flores, 2010), which allows the systematic design of images, based on the analysis of the linguistic properties of the items and the potential linguistic challenges these properties may pose to ELLs.

Procedures

Students participated individually. They were given two multiple-choice mathematics items as a warm-up to practice, so that they could familiarize themselves with three stages: 1) concurrent reporting, 2) retrospective reporting, and 3) follow-up questions asking how they solved the problems posed by the items and, in the case of illustrated items, whether and how they had used the images to solve the problems.

Each student received four items. Two items were illustrated and two items were not illustrated. One of the illustrated items and one of the non-illustrated items were the same for all students. In contrast, the other two items (one illustrated and one non-illustrated) were randomly assigned from the remaining pool of items.

Upon completing the student assent form, the researcher began recording interactions. The recording of initial interaction provides information about language use during exchanges that are not considered part of the actual interview. The recording continued throughout students' participation in concurrent reporting, retrospective reporting, and follow-up questions. All interactions were later transcribed. On average, the interview lasted about twenty minutes.

During the introduction stage of the interview, the researcher used both English and Spanish, to provide contextual clues to the participant that both languages were acceptable forms of

communication. In addition, the researcher explicitly told participants that they could speak either language (or both) as they preferred. Throughout the interview the researcher used the two languages and asked questions in both English and Spanish. This continued use of both languages was intended to provide contextual clues to students that they could use either language at any time during the interview.

Data Analysis

Consistent with the notion that each ELL has a unique pattern of L1 and L2 use, we developed a coding scheme for identifying how students respond to prompts given by the interviewer. We identified five sections of the interview to analyze language use: 1) introduction, 2) warm-up questions, 3) concurrent reporting, 4) retrospective reporting, and 5) follow-up questions. For each of these sections we coded whether the students used only Spanish, only English, or both Spanish and English. A set of rules was created to determine what counted as using one of the two languages or both. An attempt was made to code utterances that contained critical information (except for the introduction). Utterances regarded as not having critical information included speech forms such as “well” or “first.” For example, a student’s response recalling her steps that started out by saying, “first” in Spanish followed by a response completely in English was deemed not to warrant the use of two languages in the retrospective reporting.

By using this coding scheme, we are able to identify the conditions necessary to elicit comprehensive information from students about their thought processes. Additionally, we were able to compare students with similar background characteristics to identify any trends or differences in their communication when participating in the cognitive interview.

Results

Language Use

Table 1 shows the number of students who used English only and both English and Spanish during each section of the interview. Of the 31 participants, nine students used both English and Spanish in at least one section of the interview. No student used only Spanish during any of the five sections.

Table 1.

Student language use: Number of students using English only or English and Spanish

Part of the Interview	Language Used	
	English Only	English & Spanish
Introduction	25	6
Warm-Up questions	24	7
Concurrent Reporting	27	4
Retrospective Reporting	26	5
Follow-Up questions	25	6

As Table 1 shows, the majority of students used only English for most sections of the interview. 71% (22 out of 31) of the students did not use Spanish at anytime during their interview and 29% (9 out of 31) of students used Spanish at some point during the interview to convey at least one aspect of their thought processes.

Language Use and ELL Status

We examined whether the nine ELL students who used both English and Spanish during the cognitive interviews had similar backgrounds. Table 2 shows the background characteristics of ELLs using both Spanish and English during the interview. As the table shows, there was great heterogeneity in background characteristics among those students who drew from both English and Spanish during their interviews.

Table 2.

Heterogeneity in background characteristics for the nine students who used both English and Spanish during the cognitive interview.

Country of Origin	ELL Status	Years in U.S.	Age	Grade
U.S.	LEP	12	12	7
U.S.	LEP	12	12	7
U.S.	FEP	12	12	7
U.S.	LEP	14	14	8
Mexico	LEP	11	12	7
Mexico	LEP	8	13	7
Mexico	LEP	7	11	6
Mexico	NEP	2	13	8
Puerto Rico	LEP	4	13	8

Of the students who spoke both English and Spanish during the interview, four reported being born in Mexico, four reported being born in the U.S. and one reported being born in Puerto Rico. Seven students are classified as limited English proficient (LEP), one was fully English proficient (FEP), and one was non-English proficient (NEP). For those students who were born outside the U.S., the number of years the students have lived in the U.S. ranged from 2-11 years. Previous research has estimated that it can take 5-7 years for ELLs to become proficient in Academic English (Hakuta, Butler, & Witt, 2000). Based on this previous research, it might be tempting for researchers to assume that students who have lived in the country for 5 or more years not need to use Spanish because they are proficient in English. However, our results show that this is not the case.

Language Use and Background Variables

We examined language use among students with similar background variables. As shown in Table 3, Student A is classified as FEP, Students B and C are classified as LEP and Students D and E are classified as NEP. All these students are in 7th grade with the exception of Student E,

who is in 8th grade. Students A, B, and C were born in the U.S., whereas Students D and E were born in Mexico and have lived in the U.S. for less than 2 years.

As Table 3 shows, students A and B used Spanish in all parts of the interview whereas Student C communicated only in English. Even though Student A is classified as fully proficient in English, and therefore does not receive any English language support in her school day, she preferred to convey her thoughts in Spanish. During her interview she communicated almost exclusively in Spanish, especially for the retrospective reporting. Student B, classified as LEP, continually code-switched between English and Spanish throughout his entire interview. Finally, Student C, also classified as LEP, never used Spanish in any section of the interview.

Looking at the language use of Students B and C, we see the NEP students also contrasted in their language use during each section of the interview. These students are classified as non-proficient in English and receive extra English support in the form of two separate ESL classes four times a week. Student D is a 7th grader whose family regularly visits the U.S. and has currently lived in the U.S. for less than one year. Although he regularly attended English language classes in his home country of Mexico, his score on the state-mandated English development assessment classified him as not English proficient. Student E is an 8th grader who has lived in the U.S. for two years and attended school for less than one year. He is also classified as NEP on the state mandated English language proficiency exam.

Table 3

Student language use by ELL status: Selected cases.

	FEP		LEP		NEP	
	Spanish & English	English only	Spanish & English	English only	Spanish & English	English only
Introduction	1	1	1	1	0	1
Warm-up	1	1	1	1	0	1
Concurrent	1	1	1	1	0	1
Retrospective	1	1	1	1	0	1
Follow-up Questions	1	1	1	1	0	1

* FEP= Fully English proficient; LEP= Limited English proficient; NEP= Non-English proficient

As Table 3 shows, these two students' language use greatly varied. While Student D did not speak Spanish once during the interview, Student E communicated almost exclusively in Spanish. Student D showed some difficulty clearly communicating his ideas in English but preferred to use English, even when explicitly asked to use Spanish to further elaborate his thoughts. Student E was the complete opposite and communicated almost exclusively in Spanish. It was clear that Student E understood the questions in English as indicated by his responses during the concurrent, retrospective and follow-up questions. He chose to translate the majority of words to Spanish when verbalizing and elaborating his reasoning. Therefore, it appears that Student E was proficient in reading in English but preferred to speak in Spanish.

Summary and Conclusions

In this study, we addressed the need for effective approaches that take into account language variation among ELLs as critical to ensuring appropriate inclusion of these students during the process of test development. We focused on cognitive interviews as an essential component for cognitive validation. As part of a broader research project that examines the effectiveness of illustrations as a form of testing accommodations for ELLs, we conducted a series of cognitive interviews with a sample of 31 Latino, native Spanish-speaking students, classified by their school district as ELLs. We examined whether they used Spanish (L1, their native language), English (L2, their second language), or both the L1 and the L2 during each of five parts during the individual cognitive interviews: 1) introduction, 2) warm-up questions, 3) concurrent reporting, 4) retrospective reporting, and 5) follow-up questions.

We found that 26 and nine of the 31 participants respectively used only their L2 and both their L1 and L2 during the cognitive interviews. No student used only their L1 at parts of the cognitive interview other than the introduction. We also observed tremendous heterogeneity of

background variables (country of origin, ELL status, years in the U.S., age, grade) commonly used to make ELL classifications among the nine students who used both the L1 and L2 during their cognitive interviews. Moreover, we observed no correspondence between the categories of English proficiency to which these nine students were assigned and their patterns of language use during the cognitive interviews.

These results highlight the fact that including ELLs during the process of test development is absolutely necessary and that many ELLs can provide crucial information in English to researchers during the development phases. Researchers and test developers should not be discouraged from including ELLs in cognitive interviewing based on the wrong assumption that ELLs will not be able to provide useful information about the way in which they interpret test items.

However, to appropriately include ELLs, interviewers must allow students to draw from their full linguistic repertoire. This requires that interviewers themselves be fluent in both languages spoken by the student. In addition, it should be made explicit to the student that he/she can use either language at any time during the interview. Because students from various different backgrounds used both the L1 and L2, researchers must not assume or predict that any student will or will not draw from both languages. Winter, Kopriva, Chen, & Emick, (2006) are some of the few researchers who have used cognitive labs with ELLs. However, when describing language use during these investigations they report that “where possible, beginning ELL students were further paired with test administrators who spoke their native languages” (p. 271). These researchers have assumed that only beginning ELLs need to draw from both languages. However, our results show that other bilinguals, even those who are not beginning ELLs, may benefit from using both languages.

Not only did the background characteristics of students who used two languages vary, but how these students used two languages also greatly varied. Unfortunately, investigating the different types of code-switching, or use of two languages within the same conversation, is outside the scope of this project. However, we acknowledge that some students sparingly used Spanish for the purpose of accessing a word or phrase. Other students continually mixed Spanish and English throughout the entire interview. Finally, other students used Spanish during most of the interview, incorporating very little English.

In sum, our results show that broad categories used to describe background characteristics such as birthplace, length of time in the U.S., and ELL classification resulted in great heterogeneity in the language use of participants. While these variables are commonly used to make decisions about student proficiency in English, they do not provide information about how students will choose to use language during the cognitive interview. Had we used finer grain categories to describe language use, our results could have shown even more heterogeneity in the language use of participants.

Final Comments

Our study contributes to the enhancement of test development and research practices for ELLs. To our knowledge, no research has been conducted that examines how verbal protocols should be used with ELLs. In addition, very little research reports using verbal protocols with bilingual students. This study serves as a model for other studies concerning the inclusion of ELLs in large-scale testing programs.

Our results show that the majority of ELLs can provide important information through the exclusive use of English. Therefore researchers should not shy away from including bilinguals in their development phases of tests. However, in order for this inclusion to be appropriate, we

recommend that future cognitive labs conducted with ELLs always remain cognizant of the fact that bilinguals continually draw from two languages to perform tasks. Respecting this notion, we must allow these students to use either language at all times throughout the interview, regardless of background characteristics. This requires that the interviewer be fluent in the native language of the student and encourage the student to use it at anytime during the interview. It is wrong to assume that ELLs who have been classified as fully English proficient based on general background characteristics do not need to draw from both languages during cognitive interviews because. It is also wrong to assume that ELLs who have been classified as limited English proficient are incapable of providing valuable information, *in English*, on their interpretation of items.

Respecting sociolinguistic principles, we recognize that our results are drawn from a very specific interaction between the participants and the interviewer. We acknowledge the fact that using another interviewer with different characteristics (e.g., different ethnic and language background) might have yielded different results. For now, we have shown that the tremendous linguistic heterogeneity among ELLs is a fact that cannot be properly addressed based on using a few categories of language proficiency. If we are serious about including ELLs in assessment programs, we need to ensure their participation in cognitive interviews. To accomplish this, we need to develop cognitive interview procedures that are sensitive to this linguistic heterogeneity

Author's Note

The investigation reported in this paper was funded by the National Science Foundation, Grant DRK-12 DRL 0822362. The opinions expressed here are not necessarily those of the funding agency.

Notes

¹ For the purposes of this study, verbal protocols include concurrent reporting, retrospective reporting and follow-up questions. Additionally, we use the terms, *cognitive interview* and *verbal protocols* interchangeably.

² This student was classified as LEP according to the state's mandated English language acquisition. However, since she tested proficient on the state assessment in reading and writing (in English) this student did not receive extra services for English language support.

References

- Baxter, G.P., Elder, A.D., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist*, 31(2), 133-140.
- Baxter, G.P. & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice*, 17(3), 37-45.
- Bialystok, E. (2002). Cognitive processes of L2 user. In V. J. Cook (Ed.), *Portraits of the L2 user* (pp. 145–165). Clevedon, England: Multilingual Matters.
- Ericsson, K.A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. MITE Press.
- Fishman, J. A. 1965. Who speaks what language to whom and when? *La Linguistique*, 2, 67-88.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one. *Brain and Language*, 36, 3-15.
- Hamilton, L.S., Nussbaum, E.M., & Snow, R.E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, 10, 181-200.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* (policy Report 2000-1). Los Angeles, CA: University of California Linguistic Minority Research Institute.
- Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.
- Kopriva, R. (2008). *Improving testing for English language learners*. New York: Routledge.
- Moschkovich, J. (2006). Using two languages when learning mathematics. *Educational Studies in Mathematics*, 64(2), 121-144.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. In L. Wei (Ed.), *The bilingualism reader*, 221-256. New York: Routledge.
- Prosser, R., & Solano-Flores, G. (2009). *Examining the effect of rater language background on the quality of the scoring of English language learner responses to open-ended mathematics test items*. Paper presented at the annual meeting of the American Educational Research Association, April, 2009.
- Ruiz-Primo, M. A., Shavelson, R. J., Li, M., Schultz, S. E. (2001). On the validity of cognitive interpretations of scores from alternative mapping techniques. *Educational Assessment*, 7(2), 99-141.
- Solano-Flores, G., & Li, M. (2006). The use of generalizability (G) theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice* 25(1), 13-22.
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37(4), 189-199.
- Solano-Flores, G., & Li, M. (2008). Examining the dependability of academic achievement measures for English-Language Learners. *Assessment for Effective Intervention*, 33(3), 135-144.
- Solano-Flores, G., & Li, M. (2009). Generalizability of cognitive interview-based measures across cultural groups. *Educational Measurement: Issues and Practice*, 28 (2), 9-18.
- Valdes, G. & Figueroa, R.A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Albex

Winter, P.C., Kopriva, R., Chen, C. S., & Emick, J. (2006). Exploring individual and item factors that affect assessment validity for diverse learners: Results from a large-scale cognitive lab.

Learning and Individual Differences, 16, 267-276.