TEACHER ASSIGNMENT OF TEST ACCOMMODATIONS

Teacher Assignment of Test Accommodations for English Language Learners: Does it work?

Jennifer Koran, University of Maryland
Rebecca Kopriva, University of Wisconsin

Abstract

There has been a strong call for more systematic methods for selecting appropriate large scale test accommodations for students in special populations (Abedi, Courtney, & Leon, 2003; Kopriva & Mislevy, 2001; Thompson et al., 2000; Thurlow et al., 2003). Improvement in this area is especially critical for English language learners (ELLs), a group that has a relatively short history of inclusion in large scale assessments. This paper compares teacher accommodations decisions before and after a systematic data collection process and to recommendations from a multi-source, theory-driven approach for gathering information and assigning appropriate accommodations to individual ELLs. Results of the study suggest that using a structured data collection procedure with multiple sources does not have an effect on the nature of teachers' accommodations recommendations for their students and that their recommendations did not differ from a random set of accommodations which were presented. The computerized theory-driven decision process provided a significantly better fit to the systematic data collection data and may make an important contribution to the field.

Teacher Assignment of Test Accommodations for English Language Learners

Appropriate accommodations appear to be critical for the better assessment of special populations in academic accountability programs. At the individual level, when accommodation decisions are not appropriate to meet the need of the student, test results seem to often misrepresent what the student knows and can do (Kopriva, 2000; Hipolito-Delgado & Kopriva, 2006). At the aggregate level, consistent and appropriate accommodation decision making is critical to the validity of comparisons across districts and schools (Fuchs, et al., 2000; Hollenbeck et al., 1998). There has been a strong call for more systematic methods associated with selecting appropriate large scale test accommodations for students in special populations (Abedi, Courtney, & Leon, 2003; Kopriva & Mislevy, 2001; Thompson et all, 2000; Thurlow et al., 2003). One attempt to make more systematic accommodation decisions for students with disabilities used an inductive approach, systematically testing each student with different accommodations and examining which accommodations give the student a differential boost in test performance (Fuchs et al., 2005; Fuchs, et al., 2000). However, this can be extremely time consuming and may not be suitable for large-scale accommodation assignment needs. As such, other researchers have been working on developing deductive methods for matching individual students with test accommodations (Elliott, 2006; Helwig & Tindal, 2003; Kopriva, 2002; Siskind, 2004).

The research on deductive methods used to match English language learners (ELLs) with accommodations has been especially critical, as there is a shorter history of including these students in large scale assessments. Reviews of ELL accommodation policies from states and districts across the US have revealed the dramatic variability in the content and explicitness of guidelines for deciding accommodations for ELLs (e.g. Council of Chief State School Officers,

2003). An important prerequisite for the systematic application of theory in this context is to collect accurate and relevant data about the student to use as the basis for decision making. This information can then be given directly to decision makers or it can be used by a computer program for matching individual students with accommodations. It is thus critical that researchers identify salient variables and develop accurate methods for measuring these variables.

This paper introduces a multi-source, theory-driven approach for gathering information and assigning accommodations to individual ELLs and compares it to the largely current practice where decisions are generally made by a single person (commonly the student's teacher or the ELL teacher-specialist in the school). Our review of ELL test accommodations policies and practices both across the US and especially within selected states and districts confirmed that ELL teachers are the primary decision makers when it comes to deciding ELL accommodations (Council of Chief State School Officers, 2003). However, some educational agencies are beginning to have teams of people make these decisions (e.g. Texas Education Agency, 2004), similar to Individualized Education Plan (IEP) teams for students with disabilities (Elliott, 2006; Individuals with Disabilities Education Act, 1997). Such teams may be referred to as Language Proficiency Assessment Committees (LPACs), and are typically composed of persons with different perspectives and expertise, such as parents, teachers, specialists, and administrators. While efforts to bring multiple perspectives together to make accommodations decisions are commendable, researchers have noted that the group decision-making process is often not well defined or straightforward for a team of people with varying backgrounds to work effectively together to make good accommodations decisions (Swierzbin, Anderson, Spicuzza, Walz, & Thurlow, 1999). Thus, there is concern that in the absence of methods for structuring the group

decision process, teams such as LPACs may fail to fully capitalize on the potential benefits of involving multiple sources in the decision. Thus, the development of structured methods to consistently collect specific kinds of data from multiple sources and use these data in a flexible but consistent manner to make informed decisions across students, districts, and even states, presents an opportunity to improve upon current practices in making accommodations decisions for ELLs.

This paper draws upon the research and development underlying the Selection Taxonomy for English Language Learner Accommodations (STELLA), a computerized system which consists of both a structured method for collecting data from multiple informants and a theory driven system for matching appropriate accommodations to individual ELLs. Two kinds of information were compiled to inform the development of STELLA. The first type of information sought to distill salient student characteristics that are most critical for distinguishing the appropriate test accommodations to meet the individual student's need. The second activity evaluated the active characteristics of the most promising accommodations that make these conditions relevant for particular students. Figure 1 illustrates the kinds of student data selected from the literature that appear to be the most relevant for systematically assigning appropriate accommodations. Table 1 lists the preloaded accommodations that were examined and subsequently identified for this system. Once the accommodations and student variables were selected, the project worked to develop a hierarchical method of consolidating the student data and assigning accommodations to individual students (Kopriva & Koran, 2006).

The initial background research for STELLA included a survey of the ELL accommodation policies of partner states and districts and nationwide reviews, a review of the empirical literature on large scale accommodations and classroom accommodation practices with

ELLs, and a focus group study with ELL teachers. This formative process resulted in the development of three data collection forms, one for each of three critical informants: the teacher, the parent, and school records. These forms establish the basis for the systematic compilation of student information into a unified student profile. In a number of cases data collected from one source is triangulated with another. Additionally, each form collects unique information to add to the decision process, such as a parent's insights into the student's level of native language proficiency and prior and current experiences, especially as they differ from experiences of typical U.S. students. The three STELLA data collection forms were further refined through additional feedback from interviews with parents of ELLs, individual interviews with ELL teachers, a review of the data collection forms by administrators from the states and districts collaborating on STELLA, and review of the forms and decision-making algorithms by a panel of ELL experts.

STELLA has sought to be responsive to the concerns voiced by various stakeholders during the formative stage of its development. The focus group study with ELL teachers found that in following typical policy-based guidelines ELL teachers are asked to simultaneously take on competing roles as both expert decision maker and student advocate. This presents a dilemma for teachers who are "asked to make decisions with minimal guidelines and support while maintaining the best interests of the child" (Kopriva, Emick, & Hipolito-Delgado, 2006). Recommendations produced by STELLA rely directly on the insights of those who know the student personally and other extant information that may be available in the student's record. The method systematically collects information from the student's teacher, yet it also serves to separate the teacher's role as student advocate from that of primary decision maker. While STELLA collects information from the teacher about ongoing practices in the classroom, it does

not ask teachers for their recommendations. Nor does it base its decision-making algorithms on recommendations from any one source. Rather, information from multiple sources is weighed and carefully combined. In reviewing the STELLA prototype, administrators from the collaborating states and districts noted their concern that the computerization of the system would result in accommodations decisions that were cold or impersonal. However, by using an empirically based and tested system based on input from the student's teacher and others, it seems that STELLA users can be confident that the decisions retain the integrity of the connection to ongoing classroom practices and cultural complexity identified by parents and school records. A fuller discussion of STELLA can be found in Hedgespath, Koran, and Kopriva (2006) and Kopriva and Koran (2006).

This study evaluated how STELLA assignment recommendations compared to two sets of teacher recommendations and how the system functioned internally. Specifically, two research questions were investigated. First, how do teacher accommodation decisions compare with STELLA recommendations for individual students before teachers collect data on the STELLA forms and afterwards, and how do they compare with a random assignment of accommodations? Teacher accommodations before systematically collecting data on their students represented a baseline in current practice. Teachers were asked to follow the local procedures they would typically use to assign test accommodations for their students. The remaining sources of accommodations decisions represented alternatives to the status quo. Second, how are the data from the forms similar and distinct when information from the same questions was collected across more than one source, and what are the distributions of data collected within and across teachers?

Methodology

Both research questions were initially addressed by recruiting ELL teachers to take part in this study, complete questionnaires and subsequently use STELLA to collect data on their students. Further data for addressing the first research question was then produced by convening a panel of ELL experts to rate the four sets of accommodations assigned to the students.

Subjects

Data collection.

Nineteen teachers of ELLs and 114 of their students in grades kindergarten through 12th grade from the states of Maryland, North Carolina, and Texas were recruited to participate in the initial verification study. Teachers volunteered or were selected by their coordinators or supervisors to participate in the study and were compensated for their participation. Each teacher in turn recruited parents of six ELLs from his or her classes. Each parent agreed to participate in an interview with the teacher (and interpreter, as needed) and to allow the teacher to use his/her child's information in the study. Parents were compensated with a gift card for a local retail store. Tables 2 and 3 show the number of teachers from each state and their number of years teaching ELLs, respectively. Tables 4, 5 and 6 show characteristics of the students in the study: their native languages, grade levels, and the type of language program in which they are enrolled, respectively. Both students and teachers brought a variety of background characteristics to the study. Students were also varied in terms of their English and native language proficiencies, as shown in Tables 7 and 8, respectively.

Ratings.

After the data collection was completed, four ELL experts were convened to form an independent evaluation panel. Three of the panelists were teachers from the state of Maryland who had a range of classroom experience with ELLs. One was a researcher who had previous

experience as a classroom teacher and also experience with ELL testing issues in related accommodations research. These four panelists were independent from the teachers who had completed the STELLA forms during data collection.

Instruments and Procedures

Data collection.

Handbooks and other printed instructions guided the teachers through each component of the data collection process. Throughout the data collection process teachers were able to contact the research staff with any additional questions related to the research protocol. First, the teachers in the study completed questionnaires in which they were asked to make accommodation recommendations for their students using their current local protocol. This served as an ecologically valid baseline against which to compare later accommodations recommendations.

Then the teachers completed three data collection forms for each student. Figure 2 gives an overview of data collected on the Record Form, which collects information that is in the student's file at the school, such as English language proficiency test scores. Figure 3 gives an overview of data collected on the Teacher Form, which collects observations the teacher has made about the student's proficiencies, procedural skills in the classroom, and their experience with classroom accommodations. Figure 4 gives an overview of data collected on the Parent/Guardian Interview Form, which addresses the student's proficiency in the home language, prior schooling experience, familiarity with standardized testing, and other relevant experiences.

After the teachers completed the data collection forms, they completed another set of questionnaires asking them to assign test accommodations to their students on the basis of the

information contained in the three forms. It is important to note that the questionnaires completed by the teachers before and after the data collection forms were not part of the regular STELLA system. The questionnaires were developed specifically for this research study. The accommodations recommendations made by the teachers on the questionnaires had no influence on recommendations made by STELLA.

In addition to the two sets of recommendations made by the teacher, a third set of accommodations was generated by using the STELLA computerized process applied to the data collected in the forms. This process involved consolidating the information collected across the three forms into a student profile and matching this profile to an appropriate set of accommodations (Hedgespath, Koran, & Kopriva, 2006). Finally, a random set of accommodations, drawn from all the sets of accommodations proposed for any of the students in the study, was assigned to each student (Koran & Kopriva, 2006). In total, four sets of recommendations were generated in this study.

Ratings.

The four raters participated in a two to three hour long training session to orient them to the rating task. Afterwards, the four raters individually examined the three completed forms for each student, and then they rated each of the four proposed sets of accommodations for its appropriateness in meeting the individual student's test accommodation need. The raters used a seven point rating scale that ranged from completely optimal (1) to completely inappropriate (7). Raters were blind to the source of each of the proposed accommodations as well as to whether the teacher recommendation was made before or after the teacher completed the three student forms.

Research Design and Analysis

First research question: Comparison of accommodations recommendations.

First, the question of how teacher accommodation decisions compare with STELLA and random accommodations before teachers collect data on the STELLA forms and afterwards was addressed using a quantitative analysis of the ratings of these accommodations given by the rating panel and a qualitative analysis of the recommended accommodations themselves.

The ordinal nature of the seven point rating scale and differences in the use of the scale across raters necessitated the use of a model that would serve to put all ratings on a uniform interval scale. Thus, a located latent class model (Uebersax, 1993b) was fit to the rater data. This model looks at the latent structure in the data, modeling student-accommodation package pairs as if they come from a number of latent classes as well as modeling individual rater characteristics, such as rater bias and the locations of thresholds between manifest category ratings. This allowed us to carefully examine rater characteristics as well as the characteristics of the rating scale itself (Koran & Kopriva, 2006). The located latent class model estimates were then used to calculate the latent trait score on a continuous latent trait scale for each pattern of ratings. These latent trait scores were subsequently used as the dependent variable in a repeated-measures ANOVA with the source of the accommodation recommendation defining the repeated measures. Post hoc analyses were completed to further analyze significant omnibus results.

A text analysis was used to evaluate the qualitative aspects of the accommodations recommendations made by the teachers before and after completing the STELLA forms. For each student the two sets of accommodations were placed side by side and notes on the similarities and differences between the two sets of accommodations were recorded. Accommodation recommendations for students associated with the same teacher were reviewed consecutively so any patterns within a particular teacher would also be apparent. A similar

process was used to compare the accommodations recommended by the teacher after completing the STELLA forms and the accommodations recommended for the same student by STELLA.

Second research question: Comparison of information across STELLA forms.

Second, the question of how the data from the forms are similar and distinct when information from the same questions is collected across more than one source was addressed by calculating correlations on data collected across multiple informants. Teachers' ratings of students' English language proficiency (ELP) in the domains of reading, writing, speaking, and listening were compared with ELP levels provided by standardized assessments. Teachers' and parents' ratings of the student's native language (L1) proficiency in reading, writing, listening, and speaking were compared. Finally, teachers' and parents' ratings of students' familiarity with standardized tests were compared.

Results

First Research Question: Comparison of Accommodations Recommendations

First, how do teacher accommodation decisions compare with STELLA and random accommodations before teachers collect data on the STELLA forms and afterwards? The ratings on four sets of accommodations for each of the 114 students were analyzed. Thus, there were 456 cases in all (accommodation set/student combinations) for each rater to review. One rater did not rate all of the cases, so there were 124 cases (27.2%) that only had three ratings. The remaining 332 accommodation-student combinations had complete data (four ratings). The Located Latent Class Analysis program (Uebersax, 1993a) was used to fit the rater model and estimate latent trait scores on the complete data cases. Latent trait scores were obtained for the cases with missing data by fixing the latent scale to be identical to the model with the complete

cases and re-estimating the model parameters using three ratings on all 456 cases (Koran & Kopriva, 2006).

The mean latent trait score for accommodations from various sources are shown in Table 9. STELLA accommodations recommendations had the lowest mean, indicating the best fit to the student on average. The results of a repeated measures ANOVA comparing these four means are shown in Table 10. Mauchly's Test of Sphericity (W=0.813, χ^2= 23.114, df = 5, p=0.000) suggests that the data may violate the sphericity assumption. However, even with the conservative Greenhouse-Geisser correction, the omnibus F test (F(2.629, 297.069)=48.523, p=.000) was significant.

Preliminary analyses on the manifest ratings suggested that there may not be any statistical difference among the teacher (before completing forms), teacher (after completing forms), and the random conditions. Thus, a follow up repeated measures ANOVA was conducted using ratings on only the teacher and random accommodation recommendations. Results, shown in Table 11, indicate that the omnibus F test for these three conditions was not significant (again with the conservative correction due to a violation of sphericity), suggesting that there were no significant differences between the random set of accommodations and either of the teacher ratings. Nor were the teacher ratings different from one another. Any differences flagged in the first ANOVA must have been due to differences between the STELLA condition and the other conditions.

Subsequent dependent t-tests comparing STELLA to each of the conditions were conducted, using a Bonferroni-style correction to control Type I error rates. Results, shown in Table 12, were significant at the .008 level, suggesting that the STELLA condition resulted in

ratings that were significantly lower (indicating better student fit) than those for each of the three other conditions.

A qualitative review of the teacher's accommodations recommendations confirmed that teacher recommendations after completing the STELLA forms were very similar to their accommodations recommendations before completing the forms. Although the teachers collected substantial information about their students during the process of completing the forms, there was very little change their recommendations as a result. In addition, teachers tended to recommend the same kinds of accommodations for all students. This pattern was consistent throughout the data. It appears that if a teacher is familiar with a set of particular accommodations, he or she will assign that particular group of accommodations to all of the ELLs regardless of individual differences. Commonly assigned accommodations included extra time, dictionary, oral instructions, and small group administration. Further, several teachers recommended unconventional test accommodations for their ELLs. For example, one teacher recommended rewarding his or her student with a treat.

There was far less commonality between the teacher's second recommendation (after completing the forms) and the STELLA recommendation. The teacher's latter accommodation recommendations (after completing the forms) exhibited more similarity to their initial accommodation recommendations (before completing the forms) than to the STELLA accommodation recommendations.

Second Research Question: Comparison of Information across STELLA Forms

There was a high correlation between ELP ratings of teachers and those provided by assessments ($R = .853$; $R^2 = .727$), suggesting that approximately three-quarters of the variance on one of these scores can be explained by the other. A moderate correlation between teacher

and parent ratings of students' native language proficiency (R = .458; R^2 = .209) was found, suggesting that one-fifth of the variance can be predicted from the other rating. Likewise, a moderate correlation was found between teachers' and parents' ratings of students' experience with standardized testing (R = .547; R^2 = .299).

Discussion

First, the evaluation of the quality of the recommended accommodations, as defined by the expert raters, indicates that STELLA's recommendations provide a significantly better fit between characteristics of individual students and accommodations than did the teacher accommodations. In addition, teachers' recommendations before and after completing the STELLA forms were not significantly different from each other or a random set of accommodations recommended for a different student. Because the expert ratings were based on information about the student found in the data collection forms only as opposed to the broader knowledge of the student's teacher, these results could be attributed to raters having a limited picture of the student as provided in the STELLA data collection procedure. It is possible that if the data collection procedure had substantial flaws both the ratings and the STELLA recommendations would be based on this flawed information, while perhaps being a poor fit to the students themselves. Nevertheless, this explanation seems untenable because the evaluation demonstrates that both teacher accommodations were no better than random. If there would have been an important set of questions that was not asked on the forms, it seems there would have been a systematic variation between teacher recommendations and the random set. However this did not occur.

Second, the qualitative analysis suggests that the teachers in this study had a tendency to recommend the same set of accommodations to meet the needs of all six of their students even

though students differed widely on salient variables. Teachers were appropriately aware of choosing diverse students to participate in the study. Indeed, other analyses of the data collected on the forms indicate that most teachers selected a reasonably heterogeneous group of students for the study (as they were asked to). These are students that the teachers themselves recognized as having different characteristics on highly relevant variables, such as level of English language proficiency and native language proficiency (Center for the Study of Assessment Validity and Evaluation, 2005). Further, in another study teachers could reliably differentiate English language proficiency and native language proficiency for their students (Hipolito-Delgado & Kopriva, 2006). Finally, in yet another recently convened panel (Plake & Impara, 2006), disability educators displayed great expertise in identifying the needs of different students but struggled to systematically match large scale test accommodations to those needs. The results of these three studies together suggest that the assignment task (as compared to the identification task) is where the teachers are struggling to select appropriate accommodations. Perhaps this is because they have not considered the nature of the characteristics in different accommodations or because they could not systematically connect student needs with specific characteristics of different accommodations. The accumulating evidence seems to suggest that the conflicting roles of student advocate and expert accommodations decision maker found in the teacher focus group study (Kopriva, Emick, & Hipolito-Delgado, 2006) should perhaps be separated. Systems such as STELLA build upon the teacher's expertise in his or her role as student advocate while introducing psychometric analytic expertise to support their role as decision-maker.

The combined results of the qualitative and quantitative analyses of the teacher's accommodations recommendations after completing the forms and the STELLA recommendations based on the same data collection process suggests that there are important

differences in the accommodations recommended. The qualitative analysis indicated that there was a noticeable difference between the teacher's recommendation after completing the forms and the STELLA recommendation. This suggests that the second component of STELLA, the accommodation decision rules, represents a use of the data collected that is distinct from the teacher's judgment. Finally, the results indicated that a blind panel of expert raters consistently rated the STELLA recommendations as having a better fit to the student than either of the teacher recommendations or the random set of accommodations. This suggests that the STELLA decision rules may make an important contribution above and beyond the systematic data collection process. The systematic data collection process, however, is a necessary prerequisite for the use of systematic accommodation decision making.

The high correlation between ELP ratings of teachers and those provided by assessments suggests that these items on the Teacher and Records Forms are largely measuring the same skills. However, the high correlation may also have been the result of the order in which the teachers in the study completed the forms. Teachers were allowed to complete the forms in any order but were encouraged to complete the records form first so any questions left blank after looking at the student's record could be completed during the parent interview. Thus, information in the student's record may have unduly informed the teacher's judgment in completing the Teacher Form. In practice (as compared to this study) STELLA is designed so that office personnel can complete the records form. This may lower the correlation between these two sources, as teachers are asked to focus on ongoing English language proficiency as assessed informally in the classroom. Further research should continue to evaluate the level of relationship between teacher information for a student's ongoing English language proficiency in the classroom and English language proficiency tests.

The more moderate correlation between teacher and parent ratings of students' native language proficiency suggests that each source is providing some substantively different information about the same phenomena as well as providing a sufficient level of triangulation. Likewise, the moderate correlation between teachers' and parents' ratings of students' experience with standardized testing indicates that teachers and parents overlap in their judgment of the students' experience and familiarity with standardized tests, but that each also explains information that is unique. Thus, the correlations generally justify the use of multiple sources in collecting relevant data about the student's background. These results are especially supportive of the parent's role in providing information about the student to be used in making an accommodation recommendation, and of the teacher's expertise in knowing what classroom accommodations are being used and, hopefully, how students are using them.

Summary and Conclusion

The findings from this paper are an important step in developing appropriate methods that address the need for more systematic selection of large scale test accommodations for ELLs. While serving as an important prerequisite for the systematic application of theory to select appropriate accommodations, the multiple informant approach relies on persons who know the student best, structuring their insights in a way that leads to accommodations decisions that suit the needs of each student. Furthermore, it utilizes teachers' expertise in identifying student status on salient variables while not abandoning teachers to a role on the periphery of their range of specialized expertise.

Judging the appropriateness of a set of accommodations can be difficult as there is no golden standard for what an optimal accommodation recommendation is for a particular individual student. In this study, we used the teacher's recommendation following the local

accommodation decision protocol as an ecologically valid baseline point of comparison. Results suggest that using a standardized, multi-source data collection procedure alone may not improve teacher judgment in selecting accommodations in this study. Intervention in assigning proper accommodations appears to be needed as well.

Further research is warranted. A follow up study of the characteristics of students who had the poorest ratings for STELLA recommended accommodations may provide insight into areas for further accommodations matching research. It may also be helpful to complete a study that directly compares the multiple informant approach against approaches that collect data in other ways, such as by testing a student's strengths directly, not only to compare the validity of the approach, but also to look for a balance with efficiency and ease of implementation. Finally, the real test of the value of a set of accommodations may be in a study which has students take parallel forms of a test using different sets of recommended accommodations, as in Fuchs, et al (2000). The second study completed for the verification of STELLA compared scores on a math test for students assigned various accommodation packages. The study, though limited in context and student age (third and fourth grades), found that appropriate accommodations versus inappropriate accommodations appear to make a significant difference (Hipolito-Delgado & Kopriva, 2006). Other studies should look more directly at validity of scores when students take different packages of accommodations. They should also examine specific intention of assignment with regards to differences of students and within the context of different accommodations.

While a multi-source data collection procedure alone may not affect teacher accommodations decisions, it shows good potential to provide the needed input for an objective accommodations recommendation system. The results of this study support the potential of

systematic, research-based guidelines to improve the quality of the accommodations recommendation. By linking student advocates and existing records with the accommodations expertise embodied in the STELLA matching algorithm, the multi-source data collection procedure can play a vital role in improving accommodations decision making.

References

Abedi, J., Courtney, M., & Leon, S. (2003). *Research supported accommodation for English language learners in NAEP*. (CSE Technical Report 586). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Center for the Study of Assessment Validity and Evaluation. (2005). *Technical manual for the Selection Taxonomy for English Language Learner Accommodations (STELLA) system*. College Park, MD: Author.

Council of Chief State School Officers. (2003). *Annual survey of state student assessment programs: 2001-2002*. Washington, DC: Author.

Elliott, S. (2006). Selecting accommodations wisely: Facilitating test access and enhancing implementation integrity. Paper presented at Accommodating Students with Disabilities on State Assessments: What Works?, Savannah, GA, March 19-20, 2006

Kopriva, R., Emick, J.E. & Hipolito-Delgado, C.P. (2006). Do proper accommodation assignments make a difference? Examining the impact of improved decision-making on scores for ELLs. Manuscript submitted for publication.

Fuchs, L., Fuchs, D., Eaton, S.B., & Hamlett, C.B. (2005). *Dynamic Assessment of Test Accommodations*. San Antonio, TX: PsychCorp.

Fuchs, L.S., Fuchs, D., Eaton, S.B., Hamlett, C.B., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*(1), 67-81.

Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children, 69*(2), 211-225.

Hipolito-Delgado, C.P., & Kopriva, R. (2006). Assessing the Selection Taxonomy for English Language Learner Accommodations. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California.

Hollenbeck, K., Tindal, G., & Almond P. (1998). Teachers' knowledge of accommodations as a validity issue in high-stakes testing. *Journal of Special Education, 32*(3), 175-183.

Individuals with Disabilities Education Act Amendments of 1997, 20 U.S. Code, Sec. 101.

Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, D.C.: Council of Chief State School Officers.

Kopriva, R. (2002). *Taxonomy for Testing English Language Learners*. Funded proposal to the U.S. Department of Education, Grants for Enhanced Assessment Instruments Program.

Hedgespath, C., Koran, J., & Kopriva, R. (2006). Using the Selection Taxonomy for English Language Learner Accommodations (STELLA) to Make Defendable Large-scale Assessment Accommodations Decisions. Manuscript in progress.

Kopriva, R., & Koran, J. (April, 2006). *Validity of a matching system for assigning accommodations for individual English language learners*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California.

Kopriva, R. & Mislevy, R. J. (2001). *Valid Assessment of English Language Learners*. Funded proposal to the U.S. Department of Education, Institute of Education Sciences. Washington, D.C.

Koran, J. & Kopriva, R. (2006). *Placing a value on test accommodations for English language learners.* Manuscript in progress.

Plake, B., & Impara, J. (2006, March). *Accommodation Station expert panel meeting*. Savannah, GA.

Siskind, T. (2004). *Achieving Accurate Results for Diverse Learners: Accommodations and Access-Enhanced Formats for English Language Learners and Students with Disabilities.* Funded proposal to the U.S. Department of Education, Grants for Enhanced Assessment Instruments Program.

Swierzbin, B., Anderson, M.E., Spicuzza, R., Walz, L., & Thurlow, M.L. (1999). Feasibility and practicalty of a decision-making tool for standards testing of students with disabilities (Minnesota Report No. 21). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved May 2, 2005, from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/MnReport21.html

Texas Education Agency. (2004). LPAC Decision-Making Process for the Texas Assessment Program (Grades 3–12).

Thompson, S., Thurlow, M., & Walz, L. (2000). *Student perspectives on the use of accommodations on large-scale assessments* (Minnesota Report No. 35). Minneapolis, MN: University of Minnestoa, National Center on Educational Outcomes.

Thurlow, M.L., Elliott, J. L., & Ysseldyke, J.E. (2003). *Testing students with disabilities: Practical strategies for complying with district and state requirements* (2nd ed.). California: Corwin Press.

Uebersax, J.S. (1993a). LLCA: Located Latent Class Analysis Version 1.0 Program User's Manual. Winston-Salem, North Carolina: Author.

Uebersax, J.S. (1993b). Statistical modeling of expert ratings on medical treatment appropriateness. *Journal of the American Statistical Association, 88*, 421-427.

Table 1

Preloaded Accommodations

Category		Accommodation
Forms	–	Standard or some Universal Design forms
	–	Access-based form
	–	Native language or dual language forms as available
Tools	–	Bilingual word list, general or test specific
	–	Picture-word dictionary
	–	Problem solving tools
Administration	–	Oral English
	–	Oral home language
	–	Small group
	–	Individual administration
	–	Language liaison/Small group
	–	Extra time
	–	More frequent breaks
Response	–	Written in native language or code switching
	–	Oral English
	–	Oral in native language or code switching
	–	Demonstrated or modeled response

Table 2

Distribution of Teachers across States

State	Number of teachers
Maryland	4
North Carolina	4
Texas	11

Table 3

Years of Experience Teaching ELLs

Years teaching ELLs	Number of teachers
1-5	8
6-10	5
11 or more	6

Table 4

Native Languages of Students

Native language	Frequency	Percent
Arabic	4	3.5
Chinese	6	5.3
English	2	1.8
Spanish	66	57.9
Vietnamese	17	14.9
Other	19	16.7
Total	114	100.0

Table 5

Student Grade Levels

Grade	Frequency	Percent
K	1	.9
1	6	5.3
2	17	14.9
3	19	16.7
4	14	12.3
5	7	6.1
6	12	10.5
7	5	4.4
8	9	7.9
9	12	10.5
10	7	6.1
11	3	2.6
12	2	1.8
Total	114	100.0

Table 6

Type of Language Program of Students

Type of language program	Number of students
Developmental bilingual	6
Transitional bilingual	5
ESL/ESOL, contained	37
ESL/ESOL, pull-out	35
Other	31

Table 7

Frequencies of Teachers' Ratings of Students' English Language Proficiency across Domains

Level	Reading	Writing	Listening	Speaking
Beginner	31	33	21	21
Low intermediate	25	33	25	29
High intermediate	42	38	48	43
Grade level competitive	16	10	20	21
Total	114	114	114	114

Table 8

Frequencies of Parents' Ratings of Students' Native Language Proficiency across Domains

Level	Reading	Writing	Listening	Speaking
Above grade level	28	24	42	40
At grade level	35	43	51	57
Below grade level	51	47	21	17
Total	114	114	114	114

Table 9

Descriptive Statistics by Source of Accommodation Recommendation

Source of Recommendation	Mean	Standard deviation
Teacher (before)	-0.1414	1.3195
Teacher (after)	-0.0324	1.2840
STELLA	-1.6721	1.1675
Random	-0.0416	1.5170

Table 10

Repeated Measures ANOVA Results

Source	Type III sum of squares	df	Mean square	F	P-value
factor1	219.786	2.629[a]	83.603	48.523	0.000
Error(factor1)	511.839	297.069[a]	1.723		

[a] The Greenhouse-Geisser correction was used.

Table 11

Repeated Measures ANOVA with STELLA Condition Removed

Source	Type III sum of squares	df	Mean square	F	P-value
factor1	0.833	1.775[a]	0.469	0.269	0.738
Error(factor1)	349.485	200.542[a]	1.743		

[a] The Greenhouse-Geisser correction was used.

Table 12

Paired Samples T-Tests

Pair	Mean difference	95% confidence interval	t	df	P-value
Teacher (before) - STELLA	1.531	1.228 to 1.833	10.024	113	0.000
Random - STELLA	1.630	1.258 to 2.003	8.671	113	0.000
STELLA – Teacher (after)	-1.640	-1.912 to -1.368	-11.937	113	0.000

Figure Captions

Figure 1. Information collected from the student profile.

Figure 2. Information collected on the Records Form.

Figure 3. Information collected on the Teacher Form.

Figure 4. Information collected on the Parent/Guardian Interview Form.

Figure 1

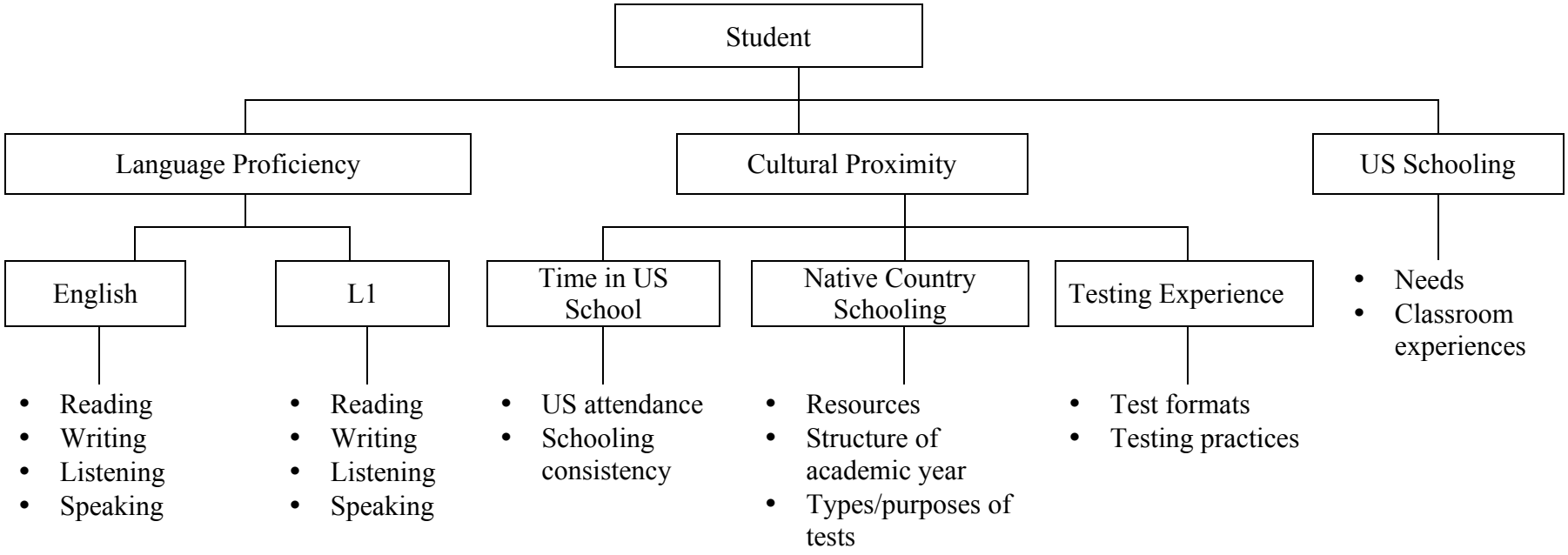
Student
Language Proficiency
Cultural Proximity
US Schooling
English
L1
Time in US School
Native Country Schooling
Testing Experience
• Needs
• Classroom experiences
• Reading
• Writing
• Listening
• Speaking
• Reading
• Writing
• Listening
• Speaking
• US attendance
• Schooling consistency
• Resources
• Structure of academic year
• Types/purposes of tests
• Test formats
• Testing practices

Figure 2

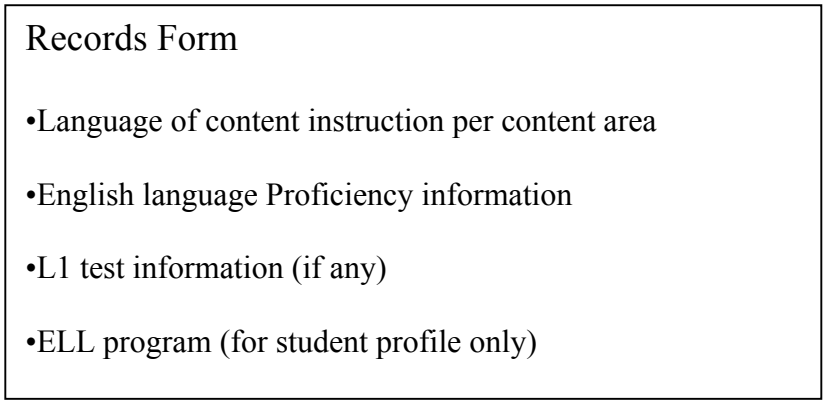
Records Form
•Language of content instruction per content area
•English language Proficiency information
•L1 test information (if any)
•ELL program (for student profile only)

Figure 3

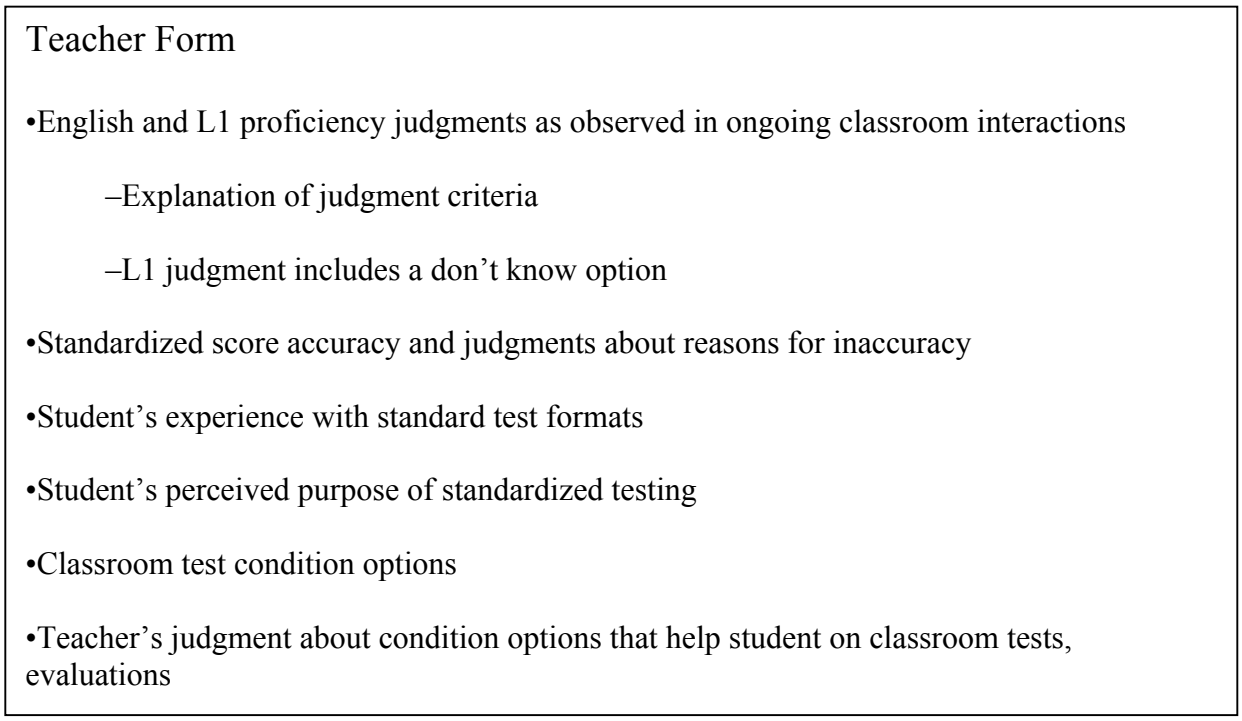
Teacher Form
•English and L1 proficiency judgments as observed in ongoing classroom interactions
–Explanation of judgment criteria
–L1 judgment includes a don't know option
•Standardized score accuracy and judgments about reasons for inaccuracy
•Student's experience with standard test formats
•Student's perceived purpose of standardized testing
•Classroom test condition options
•Teacher's judgment about condition options that help student on classroom tests, evaluations

Figure 4

Parent/Guardian Interview Form

•L1 information, 4 domains

•Full-time academic programs in U.S.

– Length of time in U.S. schools

– Consistency

•School atmosphere in native country if applicable

– Time (months, days/week, hours/day)

– Number of students in classroom

– Describing the school (e.g. chalkboards, desks, textbooks per student, other books, supplies for math or science, additional comments

•Types of tests, assessments in schools in native country

–Formal high stakes, formal not high stakes, types of ongoing classroom evaluations, methods or tasks used to assign grades

–Accurate reflection of child's achievement?

•Assessments in U.S. schools

–Accurate reflection of achievement?

–Experience with various test formats