

|| NCME MISSION FUND PROPOSAL

Authors, Affiliations, and Emails

Author	Affiliation	Email
Samuel D. Ihlenfeldt	University of Minnesota	Ihlen010@umn.edu

Descriptive Title

Using Experimental Methods to Evaluate Score Reports Adapted for English Learner Parents

Proposal Themes

Which of the following themes does your proposal address?

- ☐ Social Justice at the Center of Assessment Design
- ☒ Fairness and Equity in Validation, Reporting, and Consequences
- ☐ Innovation in Service of Fairness, Equity, and Social Justice

Purpose

In the U.S. public schools, five million English Learners (ELs) account for approximately 10% of the entire K-12 student population (USDOE, n.d.). These students face considerably more challenges at school because they must master academic content while juggling multiple languages (Sanchez, 2017). To both assess the progress of ELs and to provide meaningful feedback to the parents of these students, the U.S. federal government implemented No Child Left Behind in 2002 and, more recently, the Every Student Succeeds Act (ESSA; 2015). These policies prioritize the assessment and reporting of EL performance on measures of English language proficiency and content knowledge. ESSA stipulates that all students, including ELs, participate in yearly assessments of academic content (ESSA, 2015). For ELs, these assessments must be provided with simplified designs or appropriate accommodations to ensure that they “yield accurate data on what such students know and can do in academic content areas” (ESSA, 2015, p. 1826). Research on accommodations is robust, and every state offers a handful of

different accommodations that they believe will be uniquely helpful to the students in their state (see Rios et al., 2020 for a list of accommodations by state testing program).

The validity of inferences made from these results, however, require that parents receiving the results understand them and how they can be used. As such, ESSA stipulates that testing programs must “produce individual student interpretive, descriptive, and diagnostic reports...that allow parents...to understand and address the specific academic needs of students...in an understandable and uniform format, and to the extent practicable, in a language that parents can understand” (ESSA, 2015, p. 1827). Still, score reports for ELs have received considerably less attention than accommodations. Relatively little research has been conducted on score reporting (Hambleton & Zenisky, 2013), and much less on appropriate reporting for EL parents (Kim et al., 2016). In terms of implementation, it is uncommon for states to provide translated score reports in any language aside from Spanish, and many states don’t offer any additional materials to aid ELs and parents of ELs’ understanding of score reports (Faulkner-Bond et al., 2013; Rios & Ihlenfeldt, 2021). Digital score reports with “accommodations” such as a pop-up glossary were eagerly endorsed by small focus groups of EL parents (Kannan et al., 2018; Zapata-Rivera et al., 2014), but only a small number of states have begun to offer digitally accommodated reports (e.g., Spotlight Education).

Thus, the purpose of this research is two-fold. First, I aim to develop accommodated digital score reports that directly address the needs of EL parents by considering (a) best practices in score reporting (see Hambleton & Zenisky, 2013); (b) feedback from EL parent focus groups (e.g., Kannan et al., 2018; 2021); (c) reviews of current score reporting practices (e.g., Rios & Ihlenfeldt, 2021); (d) consultations with score reporting experts; (e) research on EL accommodations. Second, I will conduct experimental research to determine whether

accommodated score reports improve EL and non-EL parents' comprehension of—and the appropriateness of inferences made from—test scores.

Theoretical Framework

In the Standards (AERA et al. 2014), validity is defined as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA et al., 2014, p. 11). Thus, score reporting is an issue of validity: it attempts to guide stakeholders toward valid score interpretations and uses and away from misinterpretations. Because parents have little experience in the test development process, minimizing misinterpretations largely lies in the hands of developers. Although it's impossible to completely prevent misuses of test scores, making score reports more interpretable to the intended audience could mitigate that risk (O'Leary et al., 2017).

Hattie (2009) argued that collecting validity evidence for a score report was a critical part of the test validation process. He introduced a framework of principles that he believed were necessary for a score report to be considered “valid”. These were largely guided by the answers to the questions “what do you see?” and “what do you do next?”. Evidence of these principles could be derived from audience analyses. He proposed that valid score reports are a story: they need a theme (i.e., they don't present the answers to too many questions), and readers need a salient path to follow through the report with a satisfying destination at which to arrive. In practice, Van de Kleij et al. (2014) found that applying Hattie's framework when designing reports greatly increased the interpretability of their score reports.

Methodology

One shortcoming in the field of score reporting is the lack of experimental or quasi-experimental studies. Typically, published EL audience analyses are done with small focus groups (e.g., Kannan et al., 2018) or through broad reviews of practices (e.g., Faulkner-Bond et al., 2013). The conclusions from these studies are limited due to small samples or lack of a comparison group; as such, any inferences drawn about best practices might be spurious. Precedence for experimental score reporting research was demonstrated by Wainer et al. (1999) and Miller and Watkins (2010). In the present research, parents will be randomly assigned to one of two reports with or without an accommodated feature (e.g., a pop-up dual-language glossary). After they are given time to inspect the report, a survey will be administered to determine whether these alterations improve their comprehension of the content of the report and lead them to valid interpretations and uses. As I am currently a PhD student, this work will be overseen by faculty members at the University of Minnesota with experience conducting experimental research and research on ELs.

The score report development process will be guided by Hattie's (2009) theoretical framework of score report validity. The manipulated report features will be based on parent surveys (e.g., Kannan et al., 2018) and accommodation research. Mock score reports will be designed in Microsoft Publisher; they will mimic the score reports for summative assessments of academic content reviewed by Rios & Ihlenfeldt (2021).

The experiment will be developed in PsychoPy2 (Peirce et al., 2019); it will then be hosted on Pavlovia (<https://pavlovia.org/>) where it can be administered online remotely. My connections from working as an afterschool coordinator in the Minneapolis and Saint Paul public schools will be crucial for recruiting 600 participants. Three major non-English language groups

are present in Minnesota, differing in origin and primary mode of communication: Hmong, Somali, and Spanish; as such, professional translators will be hired to prepare the accommodations and questionnaires in these three languages. Individuals with a verified link will be able to access the experiment directly on Pavlovia. All participants will be compensated \$20 for approximately one hour of their time.

Anticipated Results, Outcomes, and/or Products

This research will result in a dissertation that will be presented to the University of Minnesota for the completion of a PhD in Quantitative Methods in Education. Furthermore, a manuscript will be prepared for publication, with the intent to publish in *Educational Measurement: Issues and Practice (EM:IP)*. The readers of *EM:IP* with an interest in fair score reporting practices, ELs, and parental involvement in education will likely find great value in this line of research. In addition to publishing, I aim to present the results of this research at either the 2023 or 2024 NCME annual meeting, as well as to my colleagues at Pearson and to my former coworkers at the Minnesota Department of Education. Lastly, I will promote the results of my work through posts to social media and to the NCME Graduate Student Issues Committee Facebook page.

Educational Implications of Anticipated Results

Given the experimental nature of this study, it may be possible to draw stronger conclusions surrounding the relationship between accommodated score reports and parents' understanding of those reports. As part of the completed manuscript, I will be developing a list of recommendations on score report design for EL parents; this should be highly valuable to any test developers looking to immediately improve the accessibility of their test results. The specific contents of this list will depend on the results of the study but will document: (a) which

accommodations most help and hurt parents' understanding of test scores; (b) for which parents and languages the accommodations are most beneficial; (c) common misinterpretations unique to each accommodation; (d) general parent feedback on each score report. I hope for this research to be a tool that can be referenced by test developers, similar to Hambleton & Zenisky (2013). Ultimately, the purpose of this line of research is to help parents understand how their children are learning; thus, the most ideal outcome is for parents to be better equipped to utilize test scores to help their students grow.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). <https://www.congress.gov/bill/114th-congress/senate-bill/1177>
- Faulkner-Bond, M., Shin, M., Wang, X., Zenisky, A. L., & Moyer, E. (2013, April 28–30). *Score reports for English proficiency assessments: Current practices and future directions*. Paper presented at the annual conference of the National Council on Measurement in Education, San Francisco, CA.
- Hambleton, R. K., & Zenisky, A. L. (2013). Reporting test scores in more meaningful ways: A research-based approach to score report design. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *Testing and Assessment in School Psychology and Education* (p. 479–494). American Psychological Association.

- Hattie, J. (2009). Visibly learning from reports: The validity of score reports. *Online Educational Research Journal*. <https://oerj.webspace.durham.ac.uk/wp-content/uploads/sites/249/2021/07/4.pdf>
- Kannan, P., Zapata-Rivera, D., & Bryant, A. D. (2021). Evaluating parent comprehension of measurement error information presented in score reports. *Practical Assessment, Research, and Evaluation*, 26(1), 12.
- Kannan, P., Zapata-Rivera, D., & Leibowitz, E. A. (2018). Interpretation of score reports by diverse subgroups of parents. *Educational Assessment*, 23(3), 173–194.
- Kim, A. A., Kondo, A., Blair, A., Mancilla, L., Chapman, M., & Wilmes, C. (2016). *Interpretation and use of K-12 language proficiency assessment score reports: Perspectives of educators and parents*. WIDA.
- Miller, J. A., & Watkins, M. W. (2010). The use of graphs to communicate psychoeducational test results to parents. *Journal of Applied School Psychology*, 26(1), 1–16.
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 101, Stat. 1425 (2002).
- O'Leary, T. M., Hattie, J. A., & Griffin, P. (2017). Actual interpretations and use of scores as aspects of validity. *Educational Measurement: Issues and Practice*, 36(2), 16–23.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.
- Rios, J. A., & Ihlenfeldt, S. D. (2021). State Assessment Score Reporting Practices for English Learner Parents. *Educational Measurement: Issues and Practice*, 40(3), 31–41.

- Rios, J. A., Ihlenfeldt, S. D., & Chavez, C. (2020). Are accommodations for English learners on state accountability assessments evidence-based? A multistudy systematic review and meta-analysis. *Educational Measurement: Issues and Practice*, 39(4), 65–75.
- Sanchez, C. (2017, February 23). *English language learners: How your state is doing*. nprED. <https://www.npr.org/sections/ed/2017/02/23/512451228/5-million-english-languagelearners-a-vast-pool-of-talent-at-risk>USDOE, n.d.
- U.S. Department of Education (n.d.). *Our nation’s English learners: What are their characteristics?* <https://www2.ed.gov/datastory/el-characteristics/index.html>
- Van der Kleij, F. M., Eggen, T. J., & Engelen, R. J. (2014). Towards valid score reports in the Computer Program LOVS: A redesign study. *Studies in educational evaluation*, 43, 24–39.
- Wainer, H., Hambleton, R. K., & Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement*, 36(4), 301–335.

TIMELINE AND BUDGET

Timeline for Execution

Description	Start Date	End Date	Description
Phase 1: Literature Review	January, 2022	March, 2022	The literature review is largely complete but will be reviewed by several faculty members to meet requirements for degree completion in my program.
Phase 2: Score Report and Survey Development	March, 2022	June, 2022	Development of mock score reports and survey instruments

First drafts of score reports and survey instruments	March, 2022	March, 2022	
Pilot score reports and survey instruments	March, 2022	April, 2022	Feedback will be collected from experts in assessment and score reporting
Translate score reports and survey instruments	April, 2022	May, 2022	
Final draft of score reports and survey instruments	May, 2022	June, 2022	
Phase 3: Recruitment & Administration	June, 2022	January, 2023	Recruitment and Administration will occur concurrently, pending IRB approval.
Apply for IRB approval	June, 2022	July, 2022	Application for IRB approval will be submitted through the University of Minnesota.
Recruit participants and administer experiment	July, 2022	January, 2023	
Phase 4: Write-up	January, 2023	June, 2023	The dissertation and publishable manuscript will be written concurrently.
Manuscript First Draft	January, 2023	March, 2023	
Manuscript Revisions	March, 2023	May, 2023	
Manuscript Final Draft	May, 2023	June, 2023	
Project End	June, 2023		

Estimated Budget

Amount	Use of Money	Details
\$6,000	Personnel	Hire 3 translators to translate reports, survey instruments, and participant responses. They will also transpose and translate oral responses: Hmong, 40 hours x ~\$50/hour = \$2,000 Somali, 40 hours x ~\$50/hour = \$2,000 Spanish, 40 hours x ~\$50/hour = \$2,000 The translators will be hired from Minnesota Translations (or another Minnesota agency), as they typically have specialists in these languages.
\$280	Equipment	Pavlovia credits, 1,000 x \$0.28/credit = \$280
\$12,000	Participants	Incentives for 600 participants, \$20/each

Total: \$18,280		
-----------------	--	--