Interpretation and Use of K–12 Language Proficiency Assessment Score Reports: Perspectives of Educators and Parents

WCER Working Paper No. 2016-8

November 2016

Ahyoung Alicia Kim, Akira Kondo, Alissa Blair, Lorena Mancilla, Mark Chapman, and **Carsten Wilmes**
WIDA
Wisconsin Center for Education Research
University of Wisconsin–Madison
alicia.kim@wisc.edu

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* (WCER Working Paper 2016-8). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/working-papers

Interpretation and Use of K–12 Language Proficiency Assessment Score Reports: Perspectives of Educators and Parents

Ahyoung Alicia Kim, Akira Kondo, Alissa Blair, Lorena Mancilla, Mark Chapman, and **Carsten Wilmes**

Introduction

A number of English language proficiency exams target grades K–12 English language learners (ELLs) because of the rising need to identify their needs and provide appropriate support in language learning. A good example is the WIDA ACCESS for ELLs (hereafter ACCESS), designed to measure the English language proficiency of students identified as ELLs. Every year approximately two million K–12 ELLs in the WIDA Consortium[1] take ACCESS. After students complete the exam, score reports are provided to relevant stakeholders, including teachers and parents of the students.

Because score reports are widely used by stakeholders for many purposes (e.g., placement, reclassification of ELLs), it is necessary to understand how they are interpreted and used in educational and home settings. Such information could be used to understand the usefulness of score reports and also to enhance their quality. However, there is little research on stakeholders' interpretation and use of score reports, especially in the context of K–12 ELL exams in the Unites States. Existing research is limited to teachers' interpretation of score reports (Impara, Divine, Bruce, Liverman, & Gay, 1991; Luecht, 2003; Underwood, Zapata-Rivera, & VanWinkle, 2007). For example, Impara et al. (1991) investigated the extent to which teachers were able to interpret student-level results on a standardized state assessment and the extent to which interpretive information provided on the reverse side of the student score report improved teacher understanding. Findings suggest that interpretive material helped facilitate teachers' understanding of student scores on the assessment. However, few studies have examined how stakeholders actually use the interpreted information. Moreover, very few (Miller & Watkins, 2010) have examined score reports from the parents' perspective. To gain a deeper understanding of the meaningfulness and utility of score reports, it is necessary to examine both educators' and parents' perspectives.

The study[2] on which this paper is based investigated how two stakeholder groups—K–12 ELL educators and parents—interpret and use ACCESS score reports. Findings from qualitative interviews offer implications for score report development in general and how to further enhance

---

[1] As of 2016, a total of 37 U.S. states are part of the WIDA Consortium.

[2] This study was part of a larger study that examined educators' and parents' interpretation and use of various ACCESS for ELLs score reports and resources, with the goal of further improving the score reports. The original study included gathering interview data on a prototype score report of the new ACCESS for ELLs 2.0 Individual Student Report. A new score report has been developed for ACCESS 2.0, but it has recently become available for use by educators and teachers. Therefore, this paper reports findings from the previous reports.

the quality of ACCESS score reports. In the study, we addressed the following research questions.

Research Questions

1. How do K–12 ELL educators and parents interpret the information in an English proficiency exam score report?

2. How do K–12 ELL educators and parents use the information in an English proficiency exam score report?

Literature Review

Qualities of a Good Score Report

Users must be able to easily access and interpret score reports if valid inferences are to be made based on test scores. Misinterpretations of scores potentially limit the validity of test use (AERA, APA, & NCME, 1999). According to Goodman and Hambleton (2004), good reporting of student-level results will include (1) a narrative summary of the student's assessment at the beginning of the score report, (2) information necessary to interpret assessment results, (3) a list of actions parents can take to improve their child's performance, and (4) detailed information on assessment and score results in a separate score-interpretive guide. The last two points indicate that a good score report presents the essential information of an individual learner's performance, along with actionable information that could improve her or his performance. It will also include a detailed interpretive guide.

Previous research suggests that score reports also need to be tailored to users' needs (Roberts & Gierl, 2010; Zenisky & Hambleton, 2012). This echoes Goodman and Hambleton's (2004) suggestion to make score reports personalized with useful actionable information. To better understand the users' needs, it is necessary to define the audiences and their characteristics in advance (Goodman & Hambleton, 2004; Hambleton & Zenisky, 2013; Zenisky & Hambleton, 2012).

More recent studies reflect the development of research-based score reports. For example, Zapata-Rivera and Katz (2014) introduced an approach to identify audience characteristics when designing interactive computer-based score reports. They provide examples of an audience analysis approach, which could guide the design of score reports that are interactive and technology-enhanced. They also argue that score reports should be accessible to various audiences (parents, teachers, administrators) that might lack statistical knowledge or psychometric expertise. Therefore, it is important to design score reports that minimize misinterpretation of test results by stakeholders.

In another case of research-based score report development, Zenisky, Hambleton, and Sireci (2009) examined ways to enhance the National Assessment of Educational Progress (NAEP) reporting strategies and displays of students' test performance by interviewing users and

analyzing their perspectives. They suggest that important decisions about score reports are left to the end of the test-development cycle. However, to ensure that test scores are used as intended and are viewed positively (e.g., Goodman & Hambleton, 2004), it is important to consider in advance the audiences that interpret and use such reports, and to understand what they look for. This will allow large-scale assessment results to be readable, concise, and visually attractive for a general audience (Tukey, 1990; Wainer, 1990, 1992, 1997).

Educators' and Parents' Interpretation and Use of Score Reports

Educators are one of the most important audiences of score reports. When interpreting them, teachers search for information that could be useful for instruction (Luecht, 2003; Underwood et al., 2007), such as subscores of student performance. They look for responses to questions such as, "What are my students' strengths and weaknesses?" or "What should I do to improve students' abilities?" Such information needs to be presented in the score reports with care.

However, research shows that many educators experience difficulty understanding terms in score reports, especially measurement error concepts (e.g., Lukin, Bandalos, Eckhout, & Mickelson, 2004; Zapata-Rivera, Van Winkle, & Zwick, 2010; Zapata-Rivera & Zwick, 2011; Zwick, Zapata-Rivera, & Hegarty, 2014). Users frequently misunderstood or ignored statistical jargon (e.g., standard error, significance), symbols, and technical footnotes, and some users felt intimidated (Hambleton & Slater, 1997). Therefore, empirical studies (e.g., Wainer, Hambelton, & Meara, 1999; Zwick, Zapata-Rivera, & Hegarty, 2014) explored ways to provide enhanced visual representations of statistical information so that users could draw appropriate inferences. Overall, little is known regarding how practitioners use the score reports and related documents (Zenisky & Hambleton, 2012), suggesting the need for more research in this area.

In addition to educators, parents are the primary recipients of score reports. The National Education Goals Panel (NEGP, 1998), in its guidelines for creating reports of student performance on state tests, suggest that parents need four types of information from score reports: (1) How did my child do? (2) What types of skills or knowledge does his/her performance reflect? (3) How did my child perform compared with other students in the school, district, state, and, if available, the nation? (4) What can I do to help my child improve?

Unlike teachers, parents may lack sufficient background knowledge to accurately interpret score reports. Nevertheless, as Miller and Watkins (2010) argue, it is important that parents understand test results presented in score reports; if they do not understand such reports, they might erroneously interpret placement and treatment decisions made for their children. Parents may require additional support to understand their children's performance, such as visual graphs. As with educators, there have been few examinations of how parents use the information presented in score reports, indicating a gap in the literature. Therefore, this paper focuses on the interpretation and use of score reports by educators and parents.

Methods

Participants

Both educators and parents of K–12 ELLs, who previously received an ACCESS score report, participated in the interviews. A total of 18 educators participated; they comprised 12 in-service ELL teachers and six ELL coordinators (hereafter, we use "educator" to refer to educator participants as a group). ELL coordinators all had prior experience teaching ELLs in classroom settings, but were supervising ELL teachers at the district level. The educators represented various states that use ACCESS: Alabama (*n*=1), Colorado (*n*=1), Delaware (*n*=1), Georgia (*n*=1), Illinois (*n*=1), Minnesota (*n*=2), Nevada (*n*=1), New Jersey (*n*=1), North Dakota (*n*=1), Oklahoma (*n*=1), Pennsylvania (*n*=5), Tennessee (*n*=1), and Wyoming (*n*=1). Participating educators either taught or supported ELLs ranging from Kindergarten to Grade 12. Their background information is summarized in Table 1.

***Table 1*. ELL Educators**

ID	State	Position	Grades taught/supported
AM	Tennessee	ELL teacher	5–8
CK	Pennsylvania	ELL teacher	6
CS	Minnesota	ELL teacher	K–5
DL	Colorado	ELL coordinator	PK–12
EG	Illinois	ELL teacher	6–8
IT	North Dakota	ELL teacher	9–12
JE	Georgia	ELL coordinator	PK–12
JP	Pennsylvania	ELL teacher	K–5
LG1	Pennsylvania	ELL teacher	K–12
LG2	Oklahoma	ELL coordinator	PK–12
LL	Wyoming	ELL teacher	K–6
LS1	Pennsylvania	ELL teacher	K–12
LS2	Nevada	ELL teacher	K–8
MC	New Jersey	ELL coordinator	5–8
NW	Minnesota	ELL teacher	6–11
SC	Delaware	ELL coordinator	K–12
TJ	Alabama	ELL coordinator	K–12
VU	Pennsylvania	ELL teacher	K–4

In addition, 12 parents of ELLs in Grades K–12 participated (see Table 2). Parents included both mothers (*n*=9) and fathers (*n*=3). Their first language (L1) varied widely: Arabic (*n*=1),

Japanese (*n*=2), Korean (*n*=4), Malinke (*n*=1), Mandarin (*n*=1), Spanish (*n*=2), and Urdu (*n*=1). Level of education varied as well: middle school (*n*=1), high school (*n*=1), associate's (*n*=3), bachelor's (*n*=3), master's (*n*=2), Ph.D. (*n*=2).

***Table 2*. Parents of ELLs**

ID	State	Relation to the child	Child's grade level	Home language
AT	Pennsylvania	Father	4	Malinke
AD	Wisconsin	Father	4	Spanish
HL	Wisconsin	Mother	2, 4	Korean
JL	Wisconsin	Mother	2	Korean
JY	Wisconsin	Mother	K, 2	Korean
KH	Wisconsin	Father	5	Japanese
LQ	Pennsylvania	Mother	4, 6	Arabic
MA	Wisconsin	Mother	8	Japanese
ML	Wisconsin	Mother	K, 3, and 11	Spanish
SS	Pennsylvania	Mother	5	Urdu
XC	Wisconsin	Mother	4	Mandarin
YC	Wisconsin	Mother	3	Korean

Instrument

ACCESS for ELLs Score Reports

Two ACCESS score reports were used when interviewing educators and parents: ACCESS for ELLs Teacher Report (hereafter Teacher Report) and ACCESS for ELLs Parent/Guardian Report (hereafter Parent/Guardian Report). The Teacher Report (see Appendix A) summarizes an individual ELL's performance on ACCESS and includes the following: (1) student's background information, (2) report purpose, (3) information on student's proficiency level, scale score, confidence band for each domain and composite scores, and (4) raw score and total number of items for comprehension, speaking, and writing domains. The proficiency level ranges from Level 1 (Beginner) to Level 6 (Reaching). This index is presented as a whole number (a student's language proficiency level) followed by a decimal (proportion within the proficiency level range the student's scale score represents, rounded to the nearest tenth). The proficiency levels are computed based on the scale scores, which range from 100 to 600. These are accompanied by confidence bands, which are a graphic depiction of the Standard Error of Measurement of the scale score at 95% probability.[3]

[3] For more details, refer to the WIDA ACCESS for ELLs Interpretive Guide for Score Reports (2014).

The Parent/Guardian Report (see Appendix B) provides information to parents regarding their children's performance on ACCESS. The report consists of (1) student's background information, (2) report purpose, (3) information on the student's English language proficiency level, and (4) description of the six English language proficiency levels.

Semi-structured Interview Questions

Semi-structured interview questions were prepared to obtain in-depth information on how educators and parents of K–12 ELLs interpreted and used ACCESS score reports. The interview questions were largely categorized into (1) participant's background information, (2) interpretation of score report, and (3) use of score report.

Procedures for Data Collection and Analysis

To understand educators' and parents' interpretation and use of the Teacher Report and the Parent/Guardian Report, respectively, a member of the research team interviewed participants either individually or in groups. Each individual interview took approximately 20–30 minutes. Only for parents, interviews were conducted in English or in the parents' home language by a research team member with native-like fluency in that language (if the parents had limited English proficiency). All interviews were audio recorded and transcribed for further qualitative analysis. If the parent interview was conducted in the parents' L1, the interview was transcribed first in their L1 and then translated into English for analysis.

Interview data were qualitatively analyzed a via an iterative process (Dörnyei, 2007) by adapting approaches to coding qualitative data (e.g., Hesse-Biber & Leavy, 2005). First, a content analysis was conducted to code relevant data from the interviews. Then, relevant responses were parsed and tagged using a priori and grounded codes, generating three broad themes: (1) interpretation of score report, (2) use of score report, and (3) suggestions for improvement. Two independent coders reviewed interview data; any discrepancies between them were resolved through discussion prior to further analysis.

Findings

Educators' Interpretation of ACCESS Score Reports

In general, educators found the score report information (e.g., proficiency levels, raw scores) helpful for interpreting student performance on ACCESS. However, technical terminologies (e.g., scale score, confidence band) were deemed difficult to comprehend. Thus, they had to refer to resources, such as the Score Report Interpretive Guide, to better interpret the score report.

Proficiency Level. Educators found the *proficiency level* information of the four language domains and the overall composite to be the most helpful for understanding students' test performance. For example, LG2 shared that "the first spot to get to the information is English proficiency by language domain. And then overall proficiency level." Likewise, NW stated, "You administer tests, these [proficiency levels] are the four scores that are the most important.

And that's really like the most important information that is on this score report as well as the overall score." VU describes how she interprets the language domain and the overall proficiency level information in the score report:

> For our administrative purposes in Pennsylvania, we look at the overall composite scores for AMAOs [annual measurable achievement objectives]. And also, the composite if it's above a 5, which is one of the three prongs for exiting services in Pennsylvania. I do look at the composite scores, but I like to look at each domain to know what we need to work on this school year. Most of my students, their strongest scores are listening, so that's not the area where I need to spend the bulk of my instruction on. Whereas writing is where I usually need to spend my time on with instruction. I would say it's definitely composite scores and domains scores are what I look at the most. (VU, grades K-4 ELL teacher)

In addition, the proficiency level is used to further interpret students' *growth.* CS, EG, JP, and NW shared that they used the proficiency level to check for growth. For example, CS did this by comparing the proficiency level index in the current report with that from the previous year's report. She added that "if there's a growth … a child should … at least grow .5."

Scale Scores. Growth can be measured by referring to the scale scores as well, which educators found helpful. For example, JP compared the current scale scores with those from the previous year's report to identify growth in students' English proficiency. "The two main sections for me," he stated, "would be to look at the scale scores to compare that with the previous year's data to know how the student has demonstrated growth."

However, some educators (e.g., AM, EG, LS2) found technical terms such as *scale score* or *confidence band* to be unclear. AM was unsure of the meaning of the scale score. He stated: "This is lack of my background knowledge about looking at scale scores.… I just don't know." As suggested by LS2, this might be partly attributable to the lack of training that educators receive in regards to these technical terms.

Raw Scores. In addition to the proficiency level index of language domains and the overall composite, educators found the *raw scores* to be helpful, but they exhibited mixed feelings. By definition, the raw scores represent the items students responded to correctly out of a *total possible points* in relation to each language standard measured on ACCESS (refer to the WIDA ELD Standards for more details). According to the interview findings, educators often referred to the raw scores for interpreting student performance in relation to the language standards. IT and LG1, for example, found them to be helpful because it gave them a deeper understanding of students' strengths and weaknesses. In addition, DL commented that the raw scores were particularly informative for data-driven instruction:

> And I'd like to speak as a school leader and supervisor school leader…. Something we're focusing on, and I know it's happening throughout the nation, is ELD, English language development. So what we're doing is that teachers are getting much more sophisticated in data-driven instructions…. What we're doing, for example, when we look at ELD, we

> look at the speaking tasks to make sure teachers understand, you know, what's happening with the social instruction 'cause so many teachers go by that. And we're really making sure that they understand the language arts piece, social studies, mathematics, science, other areas. (DL, grades PK-12 ELL teacher)

Moreover, some educators considered raw scores to be beneficial, especially for content teachers. LG2 said, "One of the pieces that we'll be emphasizing, especially with our secondary school general education teachers, is the breakdown of language by content."

However, other educators shared concerns regarding the raw scores. Despite their helpfulness, raw scores were confusing and misleading, especially for content teachers. For example, EG felt concerned that raw score information would confuse non-ELL teachers due to their lack of understanding of the English language proficiency levels and what they mean on ACCESS: "The speaking, like the raw scores on the speaking tasks can be kind of helpful. But it's harder for non-ELL teachers to kind of look at this.… [I]t really wouldn't be helpful for somebody who doesn't know what 4 means or whatever."

Score Report Interpretive Guide. When interpreting the score report, educators sometimes referred to resources, such as the Score Report Interpretive Guide. However, they often lacked sufficient time to carefully examine students' performance by closely referring to these resources. For instance, LL found it cumbersome to refer to the *Score Report Interpretive Guide*. "The teachers still have to go find [it]…. [T]eachers don't have the time to go … access [it]."

Educators' Use of ACCESS Score Reports

Educators reported having used score report information for various purposes. The primary purpose was to inform ELL instruction and educational decisions at the classroom level (e.g., inform instruction), school level (e.g., make programming decision), and district level (e.g., teacher training, budgeting/allocating resources). A database was often created at the school or district level to support the decision making process. In addition to helping educators make educational decisions, score reports were used to communicate with others (i.e., teachers, parents, students).

Classroom-level Decisions. Although ACCESS for ELLs is more summative in nature rather than formative, many educators reported having used ACCESS score reports for instructional purposes. CK, CS, LG1, LS2, and VU shared that they used the score report information to inform their instruction (e.g., curriculum development, lesson planning). CK reported having used ACCESS scores for curriculum design. "[I] noticed that a lot of students have a lower score in reading last year, so [I was] thinking about [making] content teaching for them to be more accessible."

At a more individual level, CS responded that she used the score report information to identify individual student's strength and weakness ("high in listening, but needs more speaking") and developed lesson plans to address the student's weakness (i.e., speaking). She

elaborated that "I can see, oh, this child is … high in listening, but needs more speaking. So … I can adjust, differentiate my lesson to make sure this person talks more."

For making educational decisions, JP and LS2 reported using *proficiency levels* the most frequently. Moreover, JE, JP, LS2, and NW referred to the *WIDA Can Do Descriptors*, which provide detailed information of what students at various levels of proficiency can do with language in school settings. JE responded that, "Teachers look specifically at these language domains and correlate them to the *Can Do Descriptors* to get some ideas for differentiation." She used the proficiency level index in conjunction with the *Can Do Descriptors*.

School-level Decisions. A number of educators (e.g., EG, LS2, VU) reported using ACCESS results for making *programming decisions* at the school level, such as grouping students into a class to differentiate instruction or to provide intervention. EG used the score report information to form classes and place students in them:

> We use it to keep track, and we have a spreadsheet of ELLs and just for programming. So I look at that a lot if there are certain classes the students need to be in, certain classes to keep them out of, or a certain team. In the summer, when we go and decide what teams to place students in, we do look at it very closely to make those programmatic decisions and for placing students. (EG, grades 6-8 ELL teacher)

Educators sometimes referred to other resources in conjunction with ACCESS for ELLs scores for making programming decisions. For instance, CS reported having used the Minnesota Comprehensive Assessment (MCA) scores, a type of statewide content exam, in addition to ACCESS for ELLs scores for placement purposes. However, CS and NW sometimes noticed a *discrepancy between the students' test scores and the information they obtained from the resources*. CS noticed that MCA score did not always correlate with ACCESS scores. That is, a high score on ACCESS did not lead to a high score on the MCA. Similarly, according to NW, ACCESS scores did not always correlate with ACT scores. Even if an ELL student was deemed proficient based on ACCESS for ELLs, she or he may not receive a "decent ACT score," which was a concern for the ELL teachers.

District-level Decisions. JE shared that ACCESS for ELLs scores were used to create district level plans, such as planning professional development, budgeting, and allocating/developing resources. Likewise, DL explained in detail how ACCESS scores are used to support professional development at the district level:

> We also look at trends, and that helps our professional development. So we find if there's some trend here we see looking at the data like, for example, comprehension of English language development. Or something specific in writing. We train about 1,000 teachers a year, every year, to work—specifically, all new teachers, all new teachers plus any others throughout the district depending on the type of professional development we're rolling out. We look at trends and based on the trends, we might have professional learning opportunities for teachers—secondary, for example in this one thing, and elementary in

> this other thing—depending on what it is so we could also use this in different levels as well as schoolwide and grade-level wide. (DL, grades PK–12 ELL coordinator)

Creating a Database. Interview findings (e.g., LG2, LS2, SC) indicate that at either the school or district level, information from an ACCESS score report is used to build an ELL database system, which ultimately supports the various decisions that educators need to make. This system allows individual teachers access to all existing data from ACCESS scores, in addition to other relevant assessment information. According to LS2, these data can be used for multiple purposes, such as monitoring student growth or sharing the information if ELL students transfer to other schools.

Likewise, LG2 described the various ways an ACCESS for ELLs database could be used, including developing individual education plans for ELLs or professional development training. As she explained, "We use that to write individual education plan for ELLs, and we use it for our … professional development training for how to understand the proficiency levels. [W]e use it in special ed analysis, we look at scale progress growth, we look at [the database] for special ed consideration."

Communicating with other Teachers, Students, and Parents. Several educators (e.g., EG, NW, JP) reported that they used the score reports to communicate with others, including other ELL teachers or content teachers, ELL students, and parents of ELLs. Often, they used the proficiency level information for communication, such as discussing the students' English proficiency level, projecting growth, or making decisions such as reclassification of ELL students. JP described this procedure in detail:

> Basically what I do is communicate information with people in my department. With the proficiency levels I'll look at the different domains to make sure students are demonstrating growth. I'll look at the overall composite scores to compare it with AMAO requirements and make sure students are making that growth for AMAO 1 and 2. And I also look at the scale scores and compare those to the graphs that show what a normal trajectory of growth looks like…. I'll share this with other members of my core team for the purposes of potentially discussing students for exit, as well as students who are making or not making adequate progress. And I'll also share this information with classroom teachers for the purposes of communicating where students are and using the *Can Do Descriptors* to plan instruction. (JP, grades K-5 ELL teacher)

As noted by EG, the information needs to be clearly discussed with content teachers for it to be useful: "If we just make copies of it and hand it out to all the teachers it's not going to be very useful without a little bit more understanding about what's on there and what it all means."

In addition, CK and LS2 spoke of how they communicated with the ELL students to discuss their performance and set goals and exit plans. Moreover, some educators used the score report to communicate with the parents during parent-teacher conferences.

Educators' Suggestions for the Future

In general, educators preferred to have more information than less for interpreting and using score report information. They requested more information regarding students' background information (e.g., home language) and test performance (e.g., growth information, test completion rate). In addition, educators wanted more enhanced resources for interpreting and using score report information, and concrete materials to support their instruction.

Details of Students' Background Information and Performance. Educators wanted more information regarding the students' background in the score report. LG1 thought having students' *home language* indicated in the score report would be helpful. Also, she added that having a translation of the score report in the students' home language automatically prepared would be beneficial.

CK, CS, and LG1 also wanted information that would help them understand students' *growth* in comparison to previous year's scores, so that they could identify if students gained proficiency since the previous year. As demonstrated by the response of CS, having previous year's scores in the score report would make it more convenient for educators to easily interpret student growth: "I think it's just a dream of mine, but it would be nice to have … the previous year, maybe…. Then we don't have to do a lot of digging … so it's just convenient for us [and] would be nice." LG1 added that growth information would not only benefit teachers but also parents.

MC also wanted information of students' *test completion rate* in case they had not fully completed the exam. Especially on a speaking test, MC said, it would be beneficial to indicate "how many [tasks] were not administered or [if the] test [was] suspended."

Enhanced Resources for Score Report Interpretation and Use. Educators also wanted *expanded Can Do Descriptors* to obtain a clearer idea of what students can demonstrate in their proficiency levels. NW wanted "more information … for reading and writing tasks, like some … more types of things that students can and cannot do…. I think that would be helpful for classroom teachers." CS, LG2, LS2, MC, and SC also wanted more information to *support their instruction*, especially for teaching content areas. However, it may not be appropriate to include such detailed-level of instructional materials in a score report of a summative test such as ACCESS. Nevertheless, educator responses suggest the need for more instructional support for teaching varying content areas. For example, LS2 wanted information on the specific lessons/tasks that she could have her students work on to improve their weaknesses.

At a broader level, some educators requested more guidance on using the WIDA ELD standards to support classroom instruction. For example, CS, an ELL teacher for Grades K–5 stated, "We would like to have more … guidance … on the standards. 'Cause we do not have the curriculum which were, I'm on that committee, we just started it to develop curriculum for ESL. Actually put down … what exactly are we going to teach."

Updated Format and Delivery of Score Reports. Some educators wanted certain information to be more clearly indicated in the score report using larger print. For instance, MC requested the composite domain information to be displayed using *larger fonts.* Educators also had suggestions regarding the *timing of receiving* the score report. They wanted reports to become available sooner. SC put it this way:

> Part of the challenge, if I could also add, is that we get this report for us probably a week and half before we start school. And so by then, you know, class lists are already made, things are already set up, and it's very difficult, especially the middle school, high school level, we try to put all of our ELLs on one team in each grade level, so they can be better supported. And we end up having to move kids around, last minute…. So, we were all hoping with the … newer tests starting earlier and being more computer-based, we might get things earlier so that a lot more prep could be handled, conversations could take place to get better use of the data earlier rather than once the kids are back. (SC, grades K-12 ELL educator)

Overall, these suggestions can help further improve the quality of ACCESS score reports and assist teachers to better interpret and use the score report information.

Parents' Interpretation of ACCESS Score Reports

Parent participants tended to direct their attention to the domain scores rather than the overall composite score, possibly because the domain scores are located at the top of the score report. However, their interpretation of the previous score report was hindered by a lack of specificity and clarity of information presented. Only about one-fourth of the parents were able to identify that Level 6 was the highest (most answered 5) and what the student's overall composite score was. This is possibly because a bar graph rather than numbers is used to display this information. Several parents made comments similar to AT, a Malinke-speaking father, who said, "I don't know if this would make sense to the kids. Oh, your speaking is 3.5. Okay, 3.5. What does that mean? Maybe it would be good to have both. The bars that name the level and the numbers." To identify what students are capable of at different proficiency levels, parents relied on common sense and previous experience rather than referencing the Description of English Language Proficiency Levels at the bottom of the page on the score report.

For most parents, their interpretation of the score report was limited because the information was not readily comprehensible or meaningful, or because translations are necessary for the information to be useful. An Arabic-speaking mother, LQ, confessed that "I didn't do anything with [the score report] because my English is bad," and she had not received a translation of the score report from the school. Although, score report translations are readily available on the WIDA website, none of the parents that we interviewed had ever received one from their children's schools.

Parents' Use of ACCESS Score Reports

Approximately one-third of the parents indicated that the previous score report was generally useful in showing where the child stands in her or his overall English language proficiency and in what areas she or he could improve. However, they found it difficult to actually use the information for assisting the child at home. HL, a Korean-speaking parent, not only shared her thoughts on the use the score reports information, but also the impracticability of carrying out these intentions consistently: "So at home, I thought about working on it together. For a while, I had them read and discuss and write, but I couldn't do it consistently." Likewise, JL, another Korean-speaking parent, said, "I'll pay more attention since that's the weak area." However, upon further inquiry, they were not able to provide details of how to do so.

Similar to the findings from the educators interviews, parents used the score reports to communicate with their children. These conversations focused on the children's progress as exemplified by the following statement from AT, a Malinke-speaking father:

> I did have a conversation with him. I explained what it means … what is his listening score, what is his writing score and things like that. And I said I will hang it here [on the wall in their home] so the next time we see it, we will compare and see how much you are making progress. (AT, parent of a 4th grade ELL)

Contrary to the findings from the teacher interviews, parents rarely used the score reports to communicate with the child's teacher. Most parent participants (9 out of 12) indicated they had not held a conversation with the child's teacher about the score report. JY, a Korean-speaking parent, stated, "The report isn't available then because it's provided towards the end of the semester. So it's not feasible to discuss it."

Parents' Suggestions for the Future

Parents would like to see other information on ACCESS score reports, including the following: (a) information about child's performance compared with other students of the same grade or native speakers, (b) explanation of Level 6, (c) ELL exit criteria, (d) definition of key terms, and (e) tips on what parents could do with children at each proficiency level. Parents also wanted some resources (e.g., interpretative guide for parents) that could enhance their interpretation and use of the score reports. For example, JY, observed, "Here it says 'academic language' and here it says 'specific academic language.' Since I don't know what they mean, all I can tell is that one is better than the other." This suggests the need for parent-friendly resources. Despite noting the additional information that would be helpful to include, a few parents cautioned against excessive length, as exemplified by JY's statement: "All parents are busy, so there's not much time to do this search online. I'd prefer to have something on the back page of the score report. Also, parents don't read if there's too much info." It is clear from this parent's quotes that there is a balance between providing helpful information that is detailed and providing so much information that parents are overwhelmed.

Parents also had a number of concerns and questions related to the score reports. These included the purpose and content of the assessment itself ("Is the test different for each age group?"), audience ("Do you compare them to native speaking children or other ELL children?"), and specific administrative and scoring procedures ("How are the scores determined?). The parents' questions also involved the potential cause and consequences of their children's test performance. SS, an Arabic-speaking mother, noticed her son's "low scores" 2 years in a row, but could not understand what had led to this. She wondered if his teacher knew more about them by stating, "I don't know what happened at school, or maybe the teacher knows more than me about why [he] got these scores." On another note, AD, a Spanish-speaking parent, wanted to know how being in a dual language program affected her child's language learning: "So, this is the disparity that exists because when taking the test, the state does not keep in mind that these children are in dual immersion. So, that is—how do I explain it?—the controversy that exists." These comments illustrate parents' concerns with their children's progress as they grapple with the potential consequences.

Discussion and Conclusion

Study findings illuminate how educators and parents interpret and use ACCESS score reports in various ways. Educators used the score report information for a wide range of purposes, from planning everyday lessons to making school-level programming decision (e.g., placement) or district-level decisions (e.g., offering professional learning opportunities). Although ACCESS is not intended for informing daily instruction, educators heavily relied on ACCESS scores due to lack of access to or familiarity with other resources to understand students' English language development. Therefore, during the interviews, educators suggested including more information in the score reports (e.g., growth information), which would better inform the educational decisions they make. However, some educators lacked understanding of certain technical terms (e.g., scale score, confidence band), which could lead to misinterpretation and poor use of the scores. Yet, due to time pressures, educators rarely referred to the WIDA Score Report Interpretive Guide, suggesting that the interpretive guide provided on the WIDA website was not used the way it was intended.

Similar to the findings from the educator interviews, parents used the information to identify their children's current proficiency level. Although some parents felt the need to address areas for improvement in their children's English proficiency, it is unclear the extent to which they were able to follow up on the idea. Most parents rarely had a chance to discuss their child's test performance with the teachers partly because the score reports were not available until the next school year. This emphasizes the need to deliver the score reports sooner to the stakeholders to maximize their intended use.

Based on feedback from educator and parent interviews, the following suggestions need to be considered for future score reports:

(1) clarify technical terms, such as scale scores and confidence band, to enhance interpretation of score report information;

(2) create resources to enhance parents' understanding of ACCESS, student performance (e.g., interpretive guide for parents), and interpretation of scores (e.g., score report translations);

(3) include information on student growth to monitor students' language development in comparison with previous years;

(4) improve the delivery timeline of score reports as currently there is a considerable gap between the time of test administration and delivery of the scores to various stakeholders; and

(5) clarify the intended purpose of ACCESS for ELLs, and what educational decisions are most appropriate to make using this information.

Due to a number of limitations of the study, the following suggestions are made for future studies. Despite recruiting efforts, relatively few from the two stakeholder groups (i.e., educators and parents) participated in the study, with fewer parents than educators (12 parents compared with 18 educators). Furthermore, the parents who did participate demonstrated higher levels of education than most parents of English learners (10 out of 12 parent participants were college educated). Another limitation was that regardless of the language of the interview or parents' literacy in English, the sample score reports during the interview were in English, and the interview protocol did not include a way of gauging parents' literacy in English. Although care was taken to interpret individual responses, caution must be exercised in drawing inferences based on study findings. Lastly, this study examined only educators and parents. Considering that ELL *students* are one of the main stakeholders of ACCESS, future studies need to investigate students' interpretation and use of the score reports.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME). (1999). *Standards for educational and psychological testing.* Washington, D.C.: Author.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Deng, N., & Yoo, H. (2009). *Resources for reporting test scores: A bibliography for the assessment community.* Bibliography prepared for the National Council on Measurement in Education. Available at http://www.ncme.org/resources/biblio1.cfm

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford, England: Oxford University Press.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education, 17*(2), 145–220. doi:10.1207/s15324818ame1702_3

Greene, J. C. (2007). *Mixed methods in social inquiry.* San Francisco, CA: Jossey-Bass.

Hambleton, R. K., & Slater, S. C. (1997). Reliability of credentialing examinations and the impact of scoring models and standard-setting policies. *Applied Measurement in Education, 10*(1), 19–28.

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.), *APA handbook of testing and assessment in psychology, Vol. 3: Testing and assessment in school psychology and education* (pp. 479–494). Washington, D.C.: American Psychological Association.

Hesse-Biber, S. N., & Leavy, P. (2005). *The practice of qualitative research*. Thousand Oaks, CA: SAGE Publications.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice, 10*, 16–18.

Knoch, U. (2012). At the intersection of language assessment and academic advising: Communicating results of a large-scale diagnostic academic English writing assessment to students and other stakeholders. *Papers in Language Testing and Assessment, 1*, 31–49. The Association for Language Testing and Assessment of Australia and New Zealand. Retrieved from http://www.altaanz.org/uploads/5/9/0/8/5908292/3_knoch.pdf

Krueger, R. A. (1998). *Analyzing and reporting focus group results.* Thousand Oaks, CA: SAGE Publications.

Luecht, R. M. (2003, April). *Applications of multidimensional diagnostic scoring for certification and licensure tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice, 23*, 26–32.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

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.

National Education Goals Panel. (1998). *Talking about tests: An idea book for state leaders.* Washington, DC: U.S. Government Printing Office.

Patton, M. Q. (2002). Qualitative research and evaluation methods. Thousand Oaks, CA: Sage.

Roberts, M., & Gierl, M. J. (2010). Developing score reports for cognitive diagnostic assessments. *Educational Measurement: Issues and Practice, 29*(3), 25–38.

Tukey, J. W. (1990). Data-based graphics: Visual display in the decades to come. *Statistical Science, 5*(3), 327-339.

Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2007). *Growing pains: Teachers using and learning to use IDMS* (Research Memorandum 08–07). Princeton, NJ: Educational Testing Service.

Wainer, H. (1990). Graphical visions from William Playfair to John Tukey. *Statistical Science, 5*(3), 340-346.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researchers, 21*(1), 14-23.

Wainer, H. (1997). *Visual revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot.* New York: Copernicus Books.

Wainer, H., Hambleton, R. K., & Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement, 36*, 301–335.

WIDA. *WIDA Can Do Descriptors, Key Uses Edition, Grades K–12.* Madison, WI: Board of Regents of the University of Wisconsin System.

WIDA (2012). *WIDA's 2012 Amplification of the English Language Development Standards, Kindergarten–Grade 12*. Madison, WI: Board of Regents of the University of Wisconsin System.

Zapata-Rivera, D., VanWinkle, W., & Zwick, R. (2010). *Exploring effective communication and appropriate use of assessment results through teacher score reports*. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

Zapata-Rivera, J. D., & Katz, I. R. (2014). Keeping your audience in mind: Applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice, 21*(4), 442–463.

Zapata-Rivera, D., & Zwick, R. (Eds.). (2011). *Improving test score reporting: Perspectives from the ETS Score Reporting Conference* (ETS Research Rep. No. 11-45). Princeton, NJ: Educational Testing Service.

Zenisky, A. L., & Hambleton, R. K. (2012). Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice, 31*(2), 21–26.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2009). Getting the message out: An evaluation of NAEP score reporting practices with implications for dissemination test results. *Applied Measurement in Education, 22*(4), 359–375.

Zwick, R., Zapata-Rivera, D., & Hegarty, M. (2014). Comparing graphical and verbal representations of measurement error in test score reports. *Educational Assessment, 19*, 116–138.

Appendix A: WIDA ACCESS for ELLs Teacher Report

WIDA CONSORTIUM

ACCESS for ELLs® English Language Proficiency Test

Teacher Report – 2013

District: Sample District			**Student:** Lastname, First 1	
School: Sample MS			**State ID:** 123456789	**District ID:**
Grade: 7	**Tier:** A	**Grade Level Cluster:** 6-8	**Birth Date:** 07/23/1999	

Report Purpose: This report provides information regarding the levels of social and academic English language proficiency the student has attained. Social language is used to communicate for everyday purposes. Academic language is used to communicate the content of language arts, mathematics, science, and social studies. This report can be used to monitor progress from year to year and to help determine instructional strategies by content areas and standards. Please refer to the ACCESS for ELLs® Interpretive Summary for more information on the meaning and use of these scores. You may also refer to the complete Interpretive Guide for Score Reports at www.wida.us for more detailed information.

Student's level of English proficiency by language domains

Language Domain	Scale Score (Possible 100 - 600)	Confidence Band (See Interpretive Summary for definitions) 100 200 300 400 500 600	Proficiency Level (Possible 1.0 - 6.0)
Listening	368	321 |---♦----| 415	4.0
Speaking	340	303 |---♦--| 377	3.0
Reading	339	317 |-♦--| 361	2.7
Writing	367	346 |-♦-| 388	3.9
Oral Language[A]	354	325 |--◊--| 383	3.7
Literacy[B]	353	339 |-◊| 367	3.4
Comprehension[C]	348	325 |-◊--| 371	3.1
Overall Score[D] (Composite)	353	340 |◊| 366	3.4

A - Oral Language = 50% Listening + 50% Speaking

B - Literacy = 50% Reading + 50% Writing

NA - Not Attempted = Student Booklet is marked with a Non-Scoring Code of Absent, Invalidate, Declined or Deferred Special Education/504

C - Comprehension = 70% Reading + 30% Listening

D - Overall Score = 35% Reading + 35% Writing + 15% Listening + 15% Speaking

Overall Scores are computed when all 4 domains have been completed

Student's performance by WIDA English Language Development Standards

Due to varying numbers of items and their levels of difficulty, raw scores should be used with caution. See the Interpretive Guide for Score Reports for details.

COMPREHENSION (Listening and Reading)

English Language Development Standards	# of Items Correct	Total # of Items
Social & Instructional Language	9	12
Language of *Language Arts*	6	9
Language of *Mathematics*	5	9
Language of *Science*	5	6
Language of *Social Studies*	6	6

SPEAKING TASKS

English Language Development Standards (Score based on # of tasks student met or exceeded)	Raw Score	Total # of Items
Social & Instructional	3	3
Language Arts/Social Studies	3	5
Mathematics/Science	2	5

NA - Not Attempted = Student Booklet is marked with a Non-Scoring Code of Absent, Invalidated, Declined or Special Education/504 Exemption

WRITING TASKS

English Language Development Standards (Scores based on writing rubric)	Linguistic Complexity		Vocabulary Usage		Language Control	
	Raw Score	Total Possible Points	Raw Score	Total Possible Points	Raw Score	Total Possible Points
Social & Instructional	3	6	2	6	3	6
Language Arts	4	6	3	6	3	6
Mathematics & Science	3	6	2	6	3	6
Language Arts & Social Studies		0		0		0

Description of Proficiency Levels

- 1 Entering – Knows and uses minimal social language and minimal academic language with visual and graphic support
- 2 Emerging – Knows and uses some social English and general academic language with visual and graphic support
- 3 Developing – Knows and uses social English and some specific academic language with visual and graphic support
- 4 Expanding – Knows and uses social English and some technical academic language
- 5 Bridging – Knows and uses social English and academic language working with grade level material
- 6 Reaching – Knows and uses social and academic language at the highest level measured by this test

Appendix B: WIDA ACCESS for ELLs Parent/Guardian Score Report

WIDA CONSORTIUM

ACCESS for ELLs® English Language Proficiency Test

Parent/Guardian Report

District:	**Student:**	
School:	**State ID:**	**District ID:**
Grade:	**Birth Date:**	

Report Purpose: This report gives information about your child's level of social and academic English language proficiency. Social language is used to communicate for everyday purposes. Academic language is used to communicate the content of language arts, mathematics, science, and social studies.

Student's English Language Proficiency Level

Test Section	1 – Entering	2 – Emerging	3 – Developing	4 – Expanding	5 – Bridging	6 – Reaching
Listening						
Speaking						
Reading						
Writing						
Oral Language [A] (Listening and Speaking						
Literacy [B] (Reading and Writing)						
Comprehension [C] (Listening and Reading)						
Overall Score [D] (Listening, Speaking, Reading, and Writing)						

Proficiency Level	Description of English Language Proficiency Levels
1 – Entering	Knows and uses minimal social language and minimal academic language with visual support
2 – Emerging	Knows and uses some social English and general academic language with visual support
3 – Developing	Knows and uses social English and some specific academic language with visual support
4 – Expanding	Knows and uses social English and some technical academic language
5 – Bridging	Knows and uses social and academic language working with grade level material
6 – Reaching	Knows and uses social and academic language at the highest level measured by this test
Other Information	Test Section is Blank – If the student was absent for this Section of the test A – Oral Language = 50% Listening + 50% Speaking – will be blank if student was absent for one or both of the Sections B – Literacy = 50% Reading + 50% Writing – will be blank if student was absent for one or both of the Sections C – Comprehension Score = 70% Reading + 30% Listening - will be blank if student was absent for one or both of the Sections D – Overall Score = 35% Reading + 35% Writing + 15% Listening + 15% Speaking - will be blank if student was absent for one or more of the Sections