

Session E-2: Validity Evidence for ELP Assessments

Panelists: *Eric Zilbert, Margaret Ho, Phoebe Winter*

Moderator: *Brenda Calderon, U.S. Department of Education, Office of State Support*

Panelists discussed the requirements in the updated Guide as they relate to ELP assessment validity. Particular attention was given to overall test design and item development and the types of validity evidence in the Guide. This session addressed critical elements 2.1 and 2.2 and section 3 of the updated Guide as applied to ELP assessment and AELPA peer review.

Ms. Margaret Ho addressed AELPA and said her comments would relate to critical elements 1, 2, 3, 4, and 6. She said schools should think about their standards and the resources available for AELPA. The reference group consists of those who are not only ELs but have significant cognitive disabilities. She said to determine the knowledge and skills that are most appropriate based on grade-level placement. There may be a process in place, but identifying the appropriate knowledge and skills can be less clear. There is an over- and under-identification issue to address. It is confounded in some places by the belief that there isn't enough information collected from families. Also, some believe ELP and significant cognitive disabilities can't be treated together.

Because there is no federal category for students with significant cognitive disabilities, the states can refer to their state categories. Concerning participation guidance for students placed in grades 3-11, the state may say it has made decisions about alternate content tests and who would be considered for AELPA. However, Ms. Ho noted that the greatest influx of EL students occurs in grades K, 1, and 2. She cautioned states to be thoughtful about how they identify for these grades. The alternate ELP is given every year. She recommended being explicit about assumptions and desired outcomes for this population by including a theory of language learning or theory of action in the front matter of the standards document. This will set the stage for what you think you are going to teach and measure. The critical elements this relates to are 2.1 and 2.2. This affects test blueprints, items, and types. In AELPA, the reference group is those who have significant cognitive disabilities as well as ELs. What do reading, writing, listening, and speaking mean for these students? You will not see the same types of behaviors as in the general population. Using AELPA, the field will push the definitions and change the constructs in these areas. How will that look so children can demonstrate what they know and can do? The domains will probably need to change as well.

There is currently a nationwide interest in identifying this population, working toward fair assessment of these students, and getting them services. The National Center on Educational Outcomes is working on publications and resources to help all these students. CCSSO is finishing an effort in which Ms. Edynn Sato participated on identifying appropriate alternate ELP standards for K-12. Since the beginning of 2018, 24 states have become involved and held workshops. They have produced descriptors and discussed standards. Participants include states with large EL populations, including Texas, which piloted the first AELPA. ELPA21 and WIDA states have been involved. A report is scheduled to come out in September 2018. Another current effort is ALTELLA's enhanced assessment grant. ALTELLA conducted on-site observations of 100 students with significant cognitive disabilities in grades K-12.

Ms. Phoebe Winter addressed element 2.1 on test design and development and element 2.2 on item development, and related evidence in these areas. She said validity is supported by a body of evidence that should include information about processes and products that occurred during test development. The underlying theory of language acquisition should drive test development for both alternate and

general assessments. She noted that validity is in everything you do. Alignment has to be addressed in several sections, although you can point back to a previous section and then explain why the information supports the current section as well. States should make these connections explicit for the reviewer.

For element 2.1, if you use a test development approach, you can show the flow of your approach. It makes it easier to collect evidence. Pay attention to accessibility and accommodation. Although this also relates to 4.2, you need to show how your underlying theory of language acquisition affected accommodation and accessibility tools, as well as item development.

There must be alignment with the four domains. A test design document must be connected to the theory of language acquisition; this is the basis of the alignment process. Good writers can tie assessment items to the theory. Keep in mind that some standards are more important than others at certain grade levels. Some alignment methodologies assume a minimal weight, but it may not be true here. The underlying theory will help you determine alignment.

If you discover that your item design doesn't align as well as hoped, explain that you did a study but it didn't work out the way you intended, so you analyzed it. Then explain how you are going to fix the problem.

Mr. Eric Zilbert addressed critical elements 3.3 and 3.4. He said the achievement-level descriptors that states write will shape what they say. In the item specifications, they have to demonstrate that ancillary constructs do not create barriers for students. It is important to have different language groups in the field test data. Items should pertain mostly to the domain being measured. Once you have domain scores, you can see how well the assessments are covering the full range of student performance. Some students have little ability in English, while others are farther along. There's a big range of performance to cover with only 20 to 22 items. You have to be careful when producing the test forms; you need to be able to measure where the students really are. We can look at the relationship between domain scores and want more analysis: exploratory studies, dimensionality studies, regression analysis, and results of other studies. Once you have scales and can report on them, you have to set cut scores. You need evidence on the reporting process used for each domain, as applied to that domain. It's not the same for all domains. The nature and make-up of the panel needs to be documented. Once you have cut scores, you will need additional information, such as 3.4, relations to other variables. In grades K, 1, and 2, if you don't have other measures from the state, consider partnering with other schools that have looked at the early grades. The map for the peer review is a good check for the technical report outline. He suggested making sure everything is included in the technical report, working with teacher judgment to triangulate cut scores, and analyzing those data.

Questions and Comments

- There's pressure to make sure tests are not too long, especially for the alternate population. The group was asked if they had thoughts on how many performance levels they should aim for on alternate assessments. One participant said at least two. Mr. Cook said he had five. He can see differentiation by level and identify the level that represents where to set the bar. Ms. Sato was looking at this also. They have three performance levels, but the effort is still in progress. She said one of those cuts separated those being considered for reclassification. Ms. Ho said the fewer categories, the better. She suggested two levels, with follow-up materials tied to the test. Then the state can say, here are some ways you can learn more about where the students are.

You need to be able to give feedback to the families and show that the students are learning something.

- Mr. Zilbert said they were providing feedback to parents and conducting a compliance exercise in a large-scale manner. They might need to hybridize it. They're working on a protocol or procedure that can be used locally to get people involved with these students. It will need to be specified in a contract, which would be hard to do later.
- Ms. Ho said there should be a feedback loop for everyone over short periods of time. Mr. Zilbert said that could provide information about large-scale assessments. There's an opportunity to think about an assessment that is not traditional.
- There is a potential interaction between disabilities that affect communication and EL learning. There are low-communication students for whom you have to decide whether or not they have attained their maximum proficiency levels. The teacher needs this information.
- Are there different standards of performance based on disability statuses? How do you individualize appropriately?
- A participant with a field perspective said that when they get students with significant cognitive disabilities, they determine whether they are ELs before developing a plan. They are concerned about the communication piece and ensuring that these students can move forward. Which system should they try to put in place first? They address this on a case-by-case basis.
- Someone suggested that ALTELLA might be able to shed light on this issue. How do they identify students as ELs, given their disabilities? A Principal Investigator from ALTELLA said one of the biggest challenges is the range of instruction in English development that is either happening or not happening for these students. Sometimes students are identified as ELs because a language other than English is spoken at home and there is no screener. They brought in English language development experts; ELs are not an easy group to define, as there is diversity among ELs. Some may be exposed to two or three languages. However, the students can always do more than expected if given instruction. Teachers need guidance on assessments and how they can move the students along in a reasonable way.
- In some cases, no instruction was happening on EL development. Children were misplaced or overidentified. Mr. Zilbert said these children need individualized benchmarks and said there must be communication with parents. He asked, "How has it gone from high chair to formal instruction?" A yardstick of some kind is needed.
- This could be thought about as an opportunity to learn. If you haven't taught them, the assessment isn't valid. Perhaps there could be a computer program of instruction for children who are not communicators. Project Core was mentioned. It is developing a program for the delivery of Universal Core vocabulary instruction and augmentative communication supports using a multi-tiered system for augmenting language (mSAL).
- There could be a set of instructional approaches that could be compared as a starting point. The issue is very complex.
- Ms. Sato asked, "Do we need to take a half step back and think about the relationship between communication competence and EL proficiency?" The group said yes.
- There's a need to bring together experts such as linguists and psychologists to figure it out.