Arizona Consequential Validity Study for the

AZELLA Kindergarten Placement Test: Final Report

Prepared by the National Center on Educational Outcomes

NATIONAL
CENTER ON
EDUCATIONAL
OUTCOMES

UNIVERSITY
OF MINNESOTA
Driven to Discover℠

NCEO Arizona Consequential Validity Project Staff

Laurene Christensen, Principal Investigator

Kristin Liu, Co-principal Investigator

Deb Albus

Elizabeth Christian

Linda Goldstone

James Hatten

Christopher Johnstone

Ross Moen

Yoko Namura

Christopher Rogers

Vitaliy Shyyan

Martha Thurlow

Jenna Ward

Yi-Chen Wu

Table of Contents

Executive Summary

During the 2012-13 school year, the National Center on Educational Outcomes (NCEO) conducted a consequential validity evaluation of the Arizona Kindergarten Placement Test (KPT) for English language learners. NCEO staff completed data gathering activities in order to address two primary evaluation questions:

1. Does the AZELLA KPT accurately place Kindergarteners into the English development services?
2. Are the intended outcomes of the placement test (i.e., getting targeted English instruction to students to more efficiently move them out of structured immersion classes) being met?

NCEO staff observed the KPT administration, conducted retrospective interviews with KPT administrators, observed the Spring AZELLA 3 administration, conducted in-depth case studies in 5 districts, facilitated an online focus group with Arizona educators, distributed a survey to all principals in schools serving Kindergarteners, and completed a data analysis examining the Spring AZELLA 3 results in light of the KPT cut scores.

Results from data collection indicated that fewer students were identified for English services. Administrators liked the format of the new test, but they thought it was too easy and many recommended the cut score be raised. In the case studies, educators generally reported that students who scored as Initial Fluent English Proficient (IFEP) were performing like their peers in the general education classrooms. An analysis of data suggested that students who were proficient on the KPT and were not proficient on the Spring AZELLA were more likely to have lower performance in the domains of reading and writing than their peers who were proficient on both tests.

Recommendations include a reconsideration of the cut score for the KPT, suggestions for revisions to the KPT, and professional development topics.

Introduction

The Arizona English Language Learner Assessment (AZELLA) is a standards-based assessment used in the state of Arizona to measure the English language proficiency of students in Kindergarten through Grade 12. The AZELLA measures skills in four domains: Listening, Speaking, Reading, and Writing. Students who indicate a language other than English in response to any of three questions on the Primary Home Language Other Than English (PHLOTE; see Appendix A) survey are administered the AZELLA Kindergarten Placement Test (KPT) so that they may be placed in the appropriate instructional setting (AZELLA Reference Manual, 2012-2013). Scores on the AZELLA KPT determine whether students meet criteria for English Language Learner (ELL) services, and are used to monitor student proficiency both while students are in ELL programming and for two years after they have exited the ELL program (AZELLA Reference Manual, 2012-2013).

The newly formatted AZELLA KPT was introduced in Arizona for the 2012-2013 school year. The decisions made based on scores from this test are high-stakes. As a result, it is crucial that the assessment's results be both valid and reliable. However, because the AZELLA KPT is new, the possible consequences of administering the test have not yet been determined. Staff from the National Center on Educational Outcomes (NCEO) partnered with the Arizona Department of Education (ADE) to conduct a multi-part evaluation examining the accuracy of placement decisions made based on the test, and whether the intended outcomes of the test are being met. In addition, the state of Arizona wished to determine whether state-prescribed test administration protocols were adhered to, because changes in protocol may impact the reliability of test scores and accuracy of decisions made on those scores.

The specific evaluation questions for this project included the following:

3. Does the AZELLA KPT accurately place Kindergarteners into the English development services?
4. Are the intended outcomes of the placement test (i.e., getting targeted English instruction to students to more efficiently move them out of structured immersion classes) being met?

NCEO undertook six activities to address these evaluation questions. The results of these activities are summarized in this report. Activity 1 consisted of observations of the administration of the AZELLA KPT followed by a short interview with the test administrator. Activity 2 consisted of retrospective interviews with school employees who had administered the AZELLA KPT in the fall of 2012. Activity 3 was an online focus group conducted with a variety of educators across the state. This activity culminated in an in-person meeting that included a consensus-building process. Activity 4 included a series of five case studies conducted in districts across the state. Activity 5 was a survey of Arizona principals to gather broad information on the administrative issues associated with the KPT. Activity 6 involved an analysis of the relationship between data from the KPT and the Spring 2013 AZELLA administration.

All NCEO staff members working on this project were required to complete and pass AZELLA KPT Administrator Training before the project began. All instruments used to collect data for the project were created by NCEO staff and were approved by the Arizona Department of Education. These instruments may be found in the appendices.

Consequential Validity Framework Underlying the Evaluation

The Arizona Department of Education asked NCEO to complete an investigation of what is sometimes referred to as "consequential validity" in the assessment literature. Messick (1989) was the first researcher to propose that an examination of a test's validity, or the degree to which the test score interpretation and use are likely to be believable (see Cronbach, 1988; Kane, 2002; Messick, 1989), include an evaluation of the consequences of an assessment. On this point specifically, Messick argued that test developers should address: "…What rationales make credible the value implications of the score interpretation and any associated implications for action; and what evidence and arguments signify the functional worth of the testing in terms of its intended and unintended consequences" (Messick, 1989, p. 5).

The concept of test validity has evolved over time from a previous focus solely on the relationship between the test purpose and the internal structure and functioning of the test, to greater inclusion of testing effects (Linn, 1997; Shepard, 1997). Shephard (1997) argued that once the concept of test validity began to include an evaluation of the accuracy of decisions made with assessment data, it became important to think about the consequences of using the assessment. This is particularly true when there are high stakes such as increased program funding for test score improvements, or when tests are used to drive changes in instruction (Camara & Lane, 2006; Shepard, 1997).

There has been debate in the test development field over whether the consequences of an assessment should be considered part of assessment validity (Crocker, 1997; Popham, 2007; Shepard, 1997). The debate centers on which party in the test development and implementation process bears the responsibility for studying, and addressing, the consequences of an assessment (Camara & Lane, 2006; Linn, 1998; Moss, 1998; Yen, 1998). Yet, there is overall agreement that it is extremely important to study those consequences, in addition to examining characteristics of the test itself (Shepard, 1997).

The Arizona Department of Education has already undertaken work to examine characteristics of the Kindergarten Proficiency Test and its functioning in relationship to other commercially available tests (Scott, 2013), along with the alignment of its content to state English proficiency standards (Christopher & Webb, 2013). NCEO's task, as an outside evaluator, was to examine the larger implications of the Kindergarten Placement Test for policy and practice in Arizona after the test had already been implemented.

This report includes a brief description of key validity, including consequential validity, concepts from the assessment literature. It also includes a discussion of how NCEO chose to study consequential validity for the AZELLA Kindergarten Placement Test.

Consequential Validity in the AERA/APA/NCME Standards

Since the early 1960s, the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) have jointly published *The Standards for Educational and Psychological Testing*. These standards act as benchmarks for evaluating tests and testing practices, although use of the standards is completely voluntary (Camera, 2010). Discussion of test validity is a key component of the standards. The current version defines test validity as: "the degree to which evidence and theory support the interpretations of test scores entailed by proposed use of tests" (AERA, APA, NCME, 1999, p. 9). The standards do not specifically refer to consequential validity by name, but as Camara and Lane (2006) point out, the recommendation to study the intended and unintended consequences of assessment is present in three separate sections of the standards. The relevant excerpts are highlighted here:

> Chapter 1 – Test Validity
>
> Standard 1.24 When unintended consequences result from test use, an attempt should be made to investigate whether such consequences arise from the test's sensitivity to characteristics other than those it is intended to assess or to the test' failure fully to represent the intended construct. (p. 23)
>
> Chapter 13 – Educational Testing and Assessment
>
> Standard 13.1 It is the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test user. (p. 145)
>
> Chapter 15 – Testing in Program Evaluation and Public Policy
>
> When educational testing programs are mandated by school, district, state, or other authorities, the ways in which test results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of tests to identify and monitor their impact and to minimize potential negative consequences. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test user. (p. 168)

According to Camara and Lane (2006), Standard 1.24 directly connects the consequences of test use to test construction, and thus, is tied to the validity of the assessment. In contrast, Standards 13.1 and 15.7 concern social policy issues that may affect choices about test use, but

that are not directly related to the validity of the assessment. NCEO's work is in keeping with Standards 13.1 and 15.7, and thus, may not affect the validity of the AZELLA Kindergarten Placement Test results. However, the findings of this work may influence policy decisions about test use, test administration, determinations of which students are placed into ELL programming, and decisions about the content of English Language Development (ELD) classrooms. In this section of the report, we use the term "consequential validity" to refer to the study of the consequences of assessment use because that is the term used by the Arizona Department of Education to refer to this project.

The Role of Consequential Validity in the Assessment and Validity Evaluation Process

Marion and Pellegrino (2006) contend that assessment validity arguments should be structured around the assessment triangle shown in Figure 1, which is a revised version of an earlier assessment triangle (Pellegrino, Chudowski, & Glaser, 2001). Figure 1 shows the triangle restructured to include the central role of consequential validity in making decisions with assessment data.

Figure 1. The Assessment Triangle and Validity Evaluation
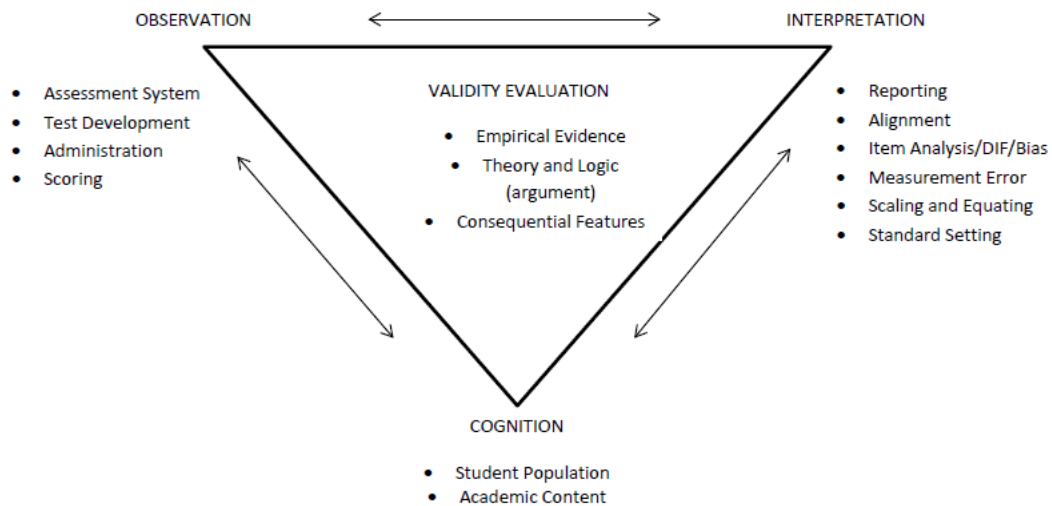
 (Marion & Pellegrino, 2006; Marion, Quenemoen, & Kearns, 2006; adapted from Pellegrino, Chudowski, & Glaser, 2001)

In Figure 1, a valid interpretation and use of test scores rests on a cyclical process beginning, in the bottom vertex, with a thorough understanding of the cognition of the population of students to be tested. An understanding of student cognition includes an understanding of how these students show what they know in a particular content area that will be assessed. This thorough understanding informs the development of test items, shown in the left vertex of the triangle. It also supports the development of appropriate test administration and scoring procedures. The

interpretation vertex, on the right, includes the various ways of sense making about, and reasoning from, students' assessment scores. An evaluation of test validity is placed at the very center of the triangle because, according to Marion (2007), "validity is clearly the most important technical criterion for the design and evaluation of assessment systems" (p. 5). In the modified assessment triangle, the validity evaluation includes an analysis of the consequences of test use. As Marion (2007) noted, implementing a new test for a special population of students will not lead to desired instructional change and improved learning outcomes for these students unless the test is accompanied by "systematic and targeted instructional interventions" (p. 18) that are either in place before the new assessment is given, or are implemented at the same time. In other words, documenting the technical quality of the assessment is not sufficient to guarantee that the test fulfills its intended purpose. An examination of the actual test use, decisions made, and consequences of the assessment, provides necessary context for evaluating test validity.

Key Points from Other Studies

NCEO examined a variety of literature recommending methods of conducting consequential validity studies. It should be noted that currently most state work on assessment validity in general, and consequential validity in particular, relates to content assessments in reading, math, and science. Work on the validity of English proficiency assessments is just beginning to emerge, in part, because the constructs underlying both the assessments and the related English proficiency standards are less clear to test developers, policymakers, and educators (Bailey, 2007; Perie & Forte, 2011). The KPT is not an accountability assessment mandated by the No Child Left Behind Act, per se. It does determine which students are classified as ELLs, receive services, and are tested with the statewide AZELLA administered at the end of each school year. Therefore, recommendations drawn from the content literature must be adapted to the particular context of the KPT. The literature suggests that:

- Consequential validity studies must address both potential positive and negative consequences of test use (Kane, 2002).
- Unintended effects of content assessments might include: (a) narrowed curriculum and instruction to only content tested; (b) intensive use of test preparation materials that are too closely tied to the test; (c) unethical test preparation; (d) differential performance gains for some subgroups; (e) inappropriate or unfair use of test scores; and (f) decreased student confidence and motivation to learn and perform well on the test (Lane & Stone, 2005).
- To date, there are no published studies of the intended or unintended consequences of a state K-12 ELL identification assessment (Perie & Forte, 2011). There are some contextualized studies of language proficiency placement test use in university language courses (e.g., Fox, 2004; Norris, 2008), but the circumstances for university assessments are quite different than for K-12 ELLs.
- Studying test consequences must take place after a new test is implemented because during development there are no consequences of test use (Reckase, 1998).

- There are no definite recommendations for how long a longitudinal study of test consequences should last (Reckase, 1998).
- Use of multiple measures is recommended to collect information on the consequences of an assessment (Lane, Park, & Stone, 1998). Typical data sources in consequential validity studies have included:
    - (a) school and district administrator surveys;
    - (b) teacher surveys;
    - (c) student surveys;
    - (d) focus groups;
    - (e) teacher interviews – either in person or via telephone (see McDonnell & Choisser, 1997);
    - (f) school administrator (and school board) interviews;
    - (g) analysis of classroom artifacts; and
    - (h) classroom observations. (Lane, Park, & Stone, 2006; Lane & Stone, 2005)
- Obtaining evidence of intended and unintended consequences from a variety of people at different levels within the school system acts as a method of confirming findings and has the potential to make the results of such a study more accurate (Lane et al., 1998).
- Studying test consequences requires a large-scale, collaborative study that may be more difficult for some parties in the test development and implementation process to coordinate than others (Green, 1998).
- Studying consequential validity is difficult work that may require substantial time, effort, and financial commitment (Linn, 1998).
- When examining potential negative consequences, evaluators must study those consequences from the perspectives of diverse stakeholders such as teachers, administrators, parents, students, etc. (Cronbach, 1989; Lane & Stone, 2005).
- Incorporating consequential validity evidence into a validity argument requires showing that the positive consequences of using a test are greater than the negative consequences (Lane & Stone, 2005).

Implications of the Literature for NCEO's Study

The studies reviewed indicated the importance of conducting an evaluation of consequential validity on as large a scale as possible in Arizona, to ensure representativeness of the results, applicability to multiple levels of educators and administrators, and to confirm findings across sources. The literature also indicated that a variety of methods can be used to collect information (e.g., focus groups, surveys, interviews), and that there is no clear guidance about the length of time for which the information should be collected. Thus, NCEO staff concluded that these details could be determined in consultation with the Arizona Department of Education. After reviewing the literature, NCEO and the Arizona Department of Education chose not to collect information from parents and students at the time of the evaluation, given the young age of the students (5-6 years old), and the fact that parents might not have experienced the consequences of the KPT first hand early in the KPT process. Instead, NCEO

chose to focus data collection efforts on the educators because they, in many cases, had administered the assessment and had experience planning instruction for students on the basis of the test results.

KPT Assessment Background

Figure 1 suggests that when determining the consequences resulting from an assessment, it is important to understand the purpose and characteristics of the assessment to be examined, who the students are and how they show what they know, and how the test data are meant to be used. In this section, we provide a brief overview of these topics before describing the consequential validity of the studies conducted.

Purpose of the assessment: Arizona Education Statute (2013) indicates that all students with a home language other than English must take an English proficiency placement test identified by the state superintendent of instruction, but does not designate which assessment must be used. As shown in the excerpt below, the purpose of the assessment is to identify children who need English language education programming:

> 15-756. Identification of English language learners
>
> A. The primary or home language for all new pupils who enroll in a school district or charter school shall be identified in a manner prescribed by the superintendent of public instruction.
> B. The English language proficiency of all pupils with a primary or home language other than English shall be assessed through the administration of English language proficiency assessments in a manner prescribed by the superintendent of public instruction. The test scores adopted by the superintendent as indicating English language proficiency shall be based on the test publishers' designated scores. The department shall annually request an appropriation to pay for the purchase of all language proficiency assessments, scoring and ancillary materials as prescribed by the department for school districts and charter schools.
> C. If it is determined that a pupil is not English language proficient, the pupil shall be classified as an English language learner and shall be enrolled in an English language education program pursuant to section 15-752 or 15-753.
>
> (Arizona Education Statute, 2013)

The students: Students who take the AZELLA Kindergarten Placement Test, or KPT, are incoming Arizona kindergartners who have been identified on a home language form, completed by parents or guardians, as speakers of a home language other than English. These students are tested to determine whether they are eligible to receive ELD services. In Arizona, the majority of the students tested during the first year were Spanish speakers, but there were students from other language backgrounds who were tested as well. These children had experienced varying degrees of native language use in the home, as well as varying degrees of familiarity with reading and writing in either the home language or English. Some of the

students had previously attended preschool and had early literacy experiences, while other students had not had these experiences at the time of testing.

The characteristics of the assessment: At the time NCEO began its work, the KPT was a new test that had been developed by Pearson Education, Inc. Ideally, it is meant to be given in the late summer, just prior to the start of the school year. However, due to a late test roll out date, in September of 2012 some districts had not yet given it at the time NCEO's work began.

The KPT is a brief (approximately 20 minute) assessment that is administered to individual students orally by a trained test administrator. Items primarily address listening, speaking, and pre-reading and pre-writing skills because new Kindergarteners generally have not yet had instruction in reading and writing at the start of the school year (Scott, 2013). At the time of testing, the test administrator scores the exam using state developed scoring criteria, and then he or she submits the score online to Pearson. The district is notified whether each student is eligible for ELD services. State communications with educators, available on the Arizona Department of Education website, indicate that those students determined to be eligible should be assigned to an ELD classroom where they receive English language instruction based on the Arizona Kindergarten English language proficiency standards. Although most schools implement a Structured English Immersion or SEI model with a 4-hour chunk of dedicated English instruction for ELLs, other English program models are possible depending on the number of ELLs in a building.

It should be noted that there is another state English proficiency assessment, the AZELLA, that students who took the KPT in the Fall of 2012 were required to take in the spring of 2013. The Spring AZELLA English Proficiency Test includes more reading and writing items, and is a longer assessment than the KPT. The results of the state AZELLA are used for school accountability purposes, to determine how many identified ELLs in grades 1-12 are making progress (called "annual measurable achievement objectives" or AMAOs) in learning English each year. Corrective action is a potential consequence for districts and schools that do not meet the AMAOs. Those students who did not pass the AZELLA are eligible for ELD services the following school year, even if they had passed the KPT in the fall. Thus, the KPT, although not a high stakes accountability test by itself, is indirectly related to high stakes for educators and schools because it partially determines which students are in the ELL subgroup for state accountability tests such as the AZELLA and other Arizona's Instrument to Measure Standards (AIMS) content assessments.

Consequential Validity Studies Conducted

The research literature on consequential validity emphasized the importance of starting assessment evaluations by describing assumptions about the test. NCEO's assumptions about the KPT were:

Assumptions about…

Giving the Test

1. The test is given to the appropriate students.
2. There is no bias introduced during the assessment process.
   a. The test administrator is well-trained in proper test administration procedures.
   b. The test is administered according to a standardized protocol and student scores consistently represent the same type of testing experience.
   c. Students understand the testing procedure.
   d. The testing environment (e.g., physical space) is appropriate for young children and does not create distractions.

Meaning of Test Scores

3. Educators, parents, and students accurately understand the purpose of the assessment.
4. High-scoring students possess the knowledge and skills in English to be successful in mainstream instruction.
5. The results of the Kindergarten placement test concur with the results of other state and district assessments, particularly the Spring AZELLA.

Use of Test Scores

6. Schools use the test scores appropriately to determine who receives language support services and who does not.
7. The adoption of the KPT will positively impact teaching and learning for both ELLs and mainstream students.

With a full understanding of the important student and assessment issues, as well as a thorough knowledge of our own assumptions about the test, NCEO staff determined that key consequential validity questions to address for this project included: (a) whether the test was being administered appropriately by trained personnel in an appropriate setting; (b) whether educators, parents, and students accurately understood the new testing policy and the purpose for implementing the new test; (c) to what degree districts and schools were using the KPT results to place students in ELD or mainstream classrooms after the school year had begun; (d) whether schools had sufficient resources (e.g., SEI teachers, teacher development opportunities, classroom space, etc.) to handle a potential change in the number of students needing ELD services as a result of implementing a new identification process; (e) whether the definition of academic language on which the test was based corresponded to the definitions of academic language held by teachers (teachers' definitions presumably influenced ELD or mainstream instructional decisions); and (f) how consistent KPT scores were with other sources of information on students' academic skills (e.g., reading assessments, curriculum-based measures, etc.).

The specific evaluation questions for NCEO's consequential validity work are presented in Table 1 along with the sources of information for each question.

Table 1. NCEO Evaluation Questions, Information Sources, and Assumptions Addressed

Questions	Information Sources	Assumption Addressed
1. Did KPT administrators follow the testing protocol?	• Test observations (Activity 1: limited number)	2b
2. Who are the KPT administrators, what kinds of training did they receive, and what are their perceptions of the testing process and the accuracy and interpretability of the results?	• Retrospective interviews with test administrators (Activity 2)	2a, 2b, 2c, 2d, 3
3. Is the test accurately placing students in ELD services or mainstream classrooms?	• Retrospective interviews with test administrators • Principal survey • Teacher interviews (Activities 2 and 4)	4, 5, 6
4. What are the implications of test results for school staffing, hiring, and training? a. What effect, if any, has implementing the assessment had on the work of specific school staff (e.g., principal, ELL coordinator, Kindergarten teachers)? b. How has staff development been affected by the implementation of the AZELLA Kindergarten placement test? c. What issues or challenges, if any, have schools experienced with school staffing and logistics as a result of the Kindergarten placement assessment?	• Principal interview – Spring 2013 • ESL coordinator interview – Spring 2013 • Principal survey – Fall 2013 (Activities 4 and 5)	7
5. What are the intended and unintended outcomes of the Kindergarten placement test?	• Focus groups with test administrators, teachers, and building administrators (Activity 3)	3, 4, 5, 6, 7
6. What are the effects, if any, of the AZELLA Kindergarten placement test on the instructional decision-making practices of schools that are/are not making Adequate Yearly Progress?	• ESL coordinator interview • Principal Interview • Kindergarten teacher interview	3, 4, 5, 6, 7

a. What types of decisions were made with AZELLA Kindergarten placement test data this year? b. What issues or challenges, if any, do schools encounter in using the test scores to make student program placement decisions? c. In what ways has the school communicated with parents about the assessment results and the related instructional decisions?	• Classroom observations (Activity 4)	
7. How consistent/inconsistent are the definitions of academic English proficiency contained in standards and assessments compared to those held by teachers and those communicated by the classroom environment? a. How does the definition of academic English proficiency held by teachers and school staff compare to the definition of academic language proficiency on which the Kindergarten placement tests (and the related standards) are based? b. How does the academic language found in the classroom environment and instruction compare/contrast with teachers views on academic language and with the assessment and the related standards?	• Conversations with state department of education staff • Document review (e.g., standards, test development materials) • Classroom observations and instructional materials review • Kindergarten teacher interviews (Activity 4)	4, 6, 7
8. How do the results of the AZELLA Kindergarten placement test compare to the results of other standardized measures of reading achievement that are given at the district or state level during the same school year? a. How do Initially Fluent English Proficient (IFEP) students score on other measures of content and English proficiency? b. How do redesignated ELLs score on other measures of content and English proficiency? c. How do current ELLs, particularly those in SEI classrooms, score on other measures of content and English proficiency?	• School/district assessment and evaluation coordinator interview • Analyses of district and school-level test scores (Activities 4 and 6)	5

Evaluation Activities

Activity 1: Observations

Observations and short follow-up interviews were conducted in four sites in Arizona to provide a description of test administration as it took place in the schools. The majority of the observations (16 of 17) were conducted between October 17 and October 19 of 2012. The final observation was conducted on November 29, 2012, when NCEO staff members were in schools conducting retrospective interviews. The employed procedures and generated findings from these observations are discussed here.

Observation Procedures

Observations and short follow-up interviews were conducted at four elementary school sites in Arizona. The sites were chosen by the Arizona Department of Education from a convenience sample of schools that had not yet completed testing at the time the current project began; still, the sites were chosen with care to include representation of urban, suburban, and charter schools. Two of the sites visited had one employee acting as a test administrator. The other two sites had multiple trained administrators (one site had 3 and one had 4). School employees administering the AZELLA KPT during observations were most often district-level employees. At one site, three district level employees were sent to conduct assessments at the school. At another site, an Assessment Director with an office at the school administered the assessment to the school's Kindergarteners. At a third site, a district-employed Language Specialist was the test administrator. At the final site, a special education teacher was trained to administer the AZELLA KPT.

NCEO staff members observed 5 test administrators giving the AZELLA KPT to 17 Kindergarten students. Each of the five educators was observed administering the AZELLA KPT to individual students. Two administrators were observed administering the test to six students; one administrator was observed administering the test to three students; and the final two administrators were observed administering the test to one student.

The majority of the students observed indicated that Spanish was their home language (see Figure 2). One student indicated a Southeast Asian home language, and one student's home language was not identified on the materials available during the administration. No Spanish was used during the administration of the AZELLA KPT; some Spanish was used by test administrators to help students understand that the test would be administered in English. No students were receiving special education services at the time of testing, but one student appeared to have speech/language difficulties.

Figure 2. Number of Students Observed, by Home Language (N=17)

Spanish (N=15)

Southeast Asian Language (N=1)

Not Identified (N=1)
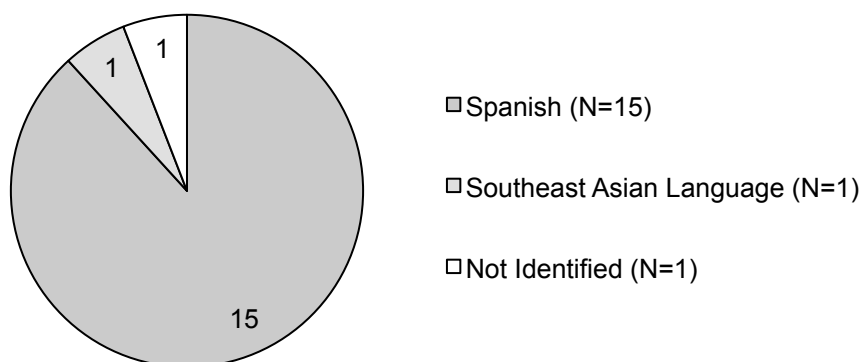
The average time to administer the test was 14.41 minutes. The fastest observed test administration was 10 minutes, and the longest observed test administration was 20 minutes. Documents provided by the Arizona Department of Education indicated that the approximate total testing time should be about 14-22 minutes. Most of the AZELLA KPT administrations observed were within this expected range of time. One student did not have a time recorded, and one student received an untimed restroom break.

Test Observation Findings

In each of the testing sessions observed, administrators were able to complete all assessment tasks with students. Notes taken by observers suggested that test administrators were able to follow the steps of the assessment, build rapport with students, and keep students engaged in the assessment tasks. Three topics in the descriptive results merit further discussion: observations of the testing environment, administration observations, and notes on score reporting and materials.

The testing environment differed among sites, but observers only noted a few issues with the environment that could potentially affect the results of the test. These issues included a student bathroom break, noise in the hallway, a second teacher entering the testing room, and an overhead light in the testing room going out and coming back on during testing. Most testing sessions (N=15) were in classrooms that appeared to be in regular use; the remaining two observations were in classrooms that did not appear to be in use. NCEO staff members noted the presence of potentially distracting items in three observations. In addition, the stickers given to students at the beginning of testing in accordance with protocol appeared to be distracting to some students. In a few testing sessions (N=5), NCEO staff noted that the furniture used by the child (the desk, the chair, or both) were inappropriately sized, creating possible discomfort or distractions for the child. All but one student took the test sitting down and facing the test materials and the administrator.

Assessment protocol was followed by test administrators in 13 of the 17 observations. Protocol was followed to a lesser degree in four observations. The nature of the changes made in the test protocol that were noted by observers included:

- Administrator did not stand test booklet upright, but laid it flat on the table in front of the student (in 1 administration)
- Administrator used positive feedback such as "Good job," "Very good," or "That's a good answer" during the test (in 4 administrations)
- Administrator over-emphasized correct choice on some first sound items (in 1 administration)
- Administrator reworded the presentation of an item (in 1 administration)

In addition, observers also noted minor incidences of student interrupting or saying something that overlapped with items being presented during test administration, often resulting in repetition by the administrator or the student (in 8 administrations).

At least one test item was repeated by administrators in 8 of the 17 testing sessions. Other events requiring the administrator to redirect the student to the test were handled efficiently. Most students responded to prompts in English only (N=13). Three students responded in English and Spanish, and information on the language used in responses was not noted on one observation form. Students responded off-topic in a few instances, but test administrators were quick to help them re-focus on the test.

Researchers from NCEO observed administrators entering student scores into the computer in a few of the testing sessions (4 of 17). The test administrator who had assessed the student entered the student's scores in three of these cases; the lead test administrator entered scores in the remaining case.

Post-Observation Interviews

Test administrators were asked several questions about training, numbers of assessments given, scoring, challenges during administration, and how they thought the administration of the assessment went for the student they had just assessed (see Appendix B). Time constraints prevented administrators from responding to every question. Because the number of responses to each question is very small, it is crucial to interpret these comments with caution.

Test administrators had completed training at different times. Some administrators had been trained several months prior to participating in the observations and interviews, and some had been trained only a week prior. Administrators indicated that administration and scoring of the assessment were not very difficult, but it could be difficult to keep some students on task. Administrators reported using different strategies for scoring the protocols of students with speech or language issues. One administrator made no adjustments, while another made an accommodation by giving credit when students knew the word in question.

Administrators provided some discussion about which items they perceived to be the easiest and the most difficult for students. Most test administrators indicated that the easiest items were those that required students to answer questions about themselves (i.e., asking their name) or identify objects. Different administrators indicated that different items presented difficulties for students. One administrator noted that it was difficult for students to name initial sounds, another mentioned that it was difficult for students to draw an object, and a third mentioned that the students had trouble with parts of a kinesthetic-related item. Two administrators noted that the final item on the test presented a challenge for students because they did not have much practice with the skill being assessed early in the school year; the skill being assessed was content-related reason sequencing.

Administrators provided a few comments about the test materials. One administrator noted that in one item, pictures may be interpreted differently so that more than one answer may be correct. In addition, a student was observed to be distracted by a potential inconsistency in the details of the pictures in a story. Other comments made by individual administrators indicated different aspects of the test that they found challenging. As a result, a few administrators admitted to presenting the items in a way that they viewed to be "less awkward," or more likely to hold the student's attention. One administrator compared the old assessment to the new one, noting that the new assessment seemed easier, did not contain reading or writing tasks, and was quicker to administer.

Activity 2: Retrospective Interviews

Retrospective interviews were conducted in 24 sites in Arizona to complement the data collected in the observations. Thirty-four retrospective interviews were conducted between November 13 and December 7 of 2012. The employed procedures and generated findings from these interviews are discussed here.

Interview Procedures

The Arizona Department of Education recruited school sites via email and phone contact. To maintain a level of confidentiality for the interviews, a single staff person at ADE made the contacts and did not share interviewee information with other members of the department to support the confidential nature of the interviews. Interviewees were told that ADE had helped arrange the interviews but that NCEO staff would not name individuals or districts in its reports to the state. The geographical locations of the sites participating in the retrospective interviews represented rural, suburban, and urban communities that were widely distributed across the state.

The number of interview participants totaled 78, with interview group sizes ranging from 1 to 6 as follows: there were 15 interviews with one person, 5 interviews with two people, 8 interviews with three people, 3 interviews with four people, 1 interview with five people, and 2 interviews with six people (1 interview is an estimate). Of the 78 people interviewed, nearly all (except 3) administered assessments; the 3 who did not directly administer the KPT were

school or district leaders who had comments to contribute about testing. The employment roles of the interviewees included Kindergarten teachers (either past or present), teachers at other grade levels, principals, assistant principals, ELL directors or coordinators, paraprofessionals, interventionists, and language specialists.

Interviewees provided estimates of the number of students who were assessed using the AZELLA KPT in the Fall of 2012. Estimates for the numbers of assessments conducted at a single school site ranged from 0 to 282, with an average of 141 assessments conducted at individual school sites. When reporting on the number of assessments that were given within a district, estimates ranged from a few hundred to 1200. Figure 3 illustrates the relationship between the sites represented in the data and the number of students reportedly assessed with the AZELLA KPT.

Figure 3. Number of Assessments Conducted by School Site[a]
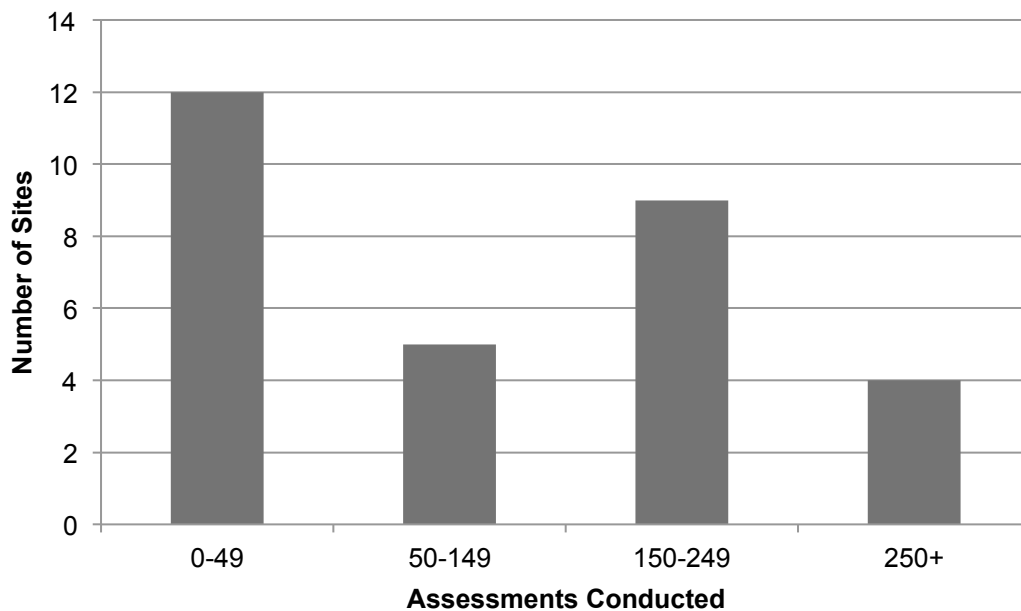
[a] n=24 schools

Though the average number of assessments conducted at an individual school site was 141, it is clear from Figure 3 that this number is influenced by the few schools that administered a large number of AZELLA KPTs. Most of the schools visited tested fewer than 250 students during the fall of 2012. It should be noted that these numbers were estimates provided by the test administrators who were interviewed; this may not accurately represent the actual number of assessments conducted at each school. In addition, test administrators in two of the interviews did not give a numerical representation of the assessments conducted, but instead indicated that "hundreds" or "a large population" were tested. Also, some interviewees held district level positions and reported on the number of assessments given in the district. These estimates are not included in Figure 3.

Summary of Retrospective Interview Findings

This section highlights the retrospective interview results, which include syntheses of educators' perceptions of AZELLA KPT administration, participating students, and test items. Other test-related insights from educators are presented as well.

Purpose of the AZELLA KPT. The interview results pointed to some dissonance in administrators' understanding of the purpose of the AZELLA KPT. Many educators perceived the test as a screener that identifies students with ELL needs. Still, a significant number of respondents described the AZELLA KPT as an English language proficiency test and pointed to the need for a stronger academic language emphasis and more prescriptive performance levels. While most descriptions of the purpose of the test were general in nature, some administrators were more specific in their perception of the role of the test: "to assess oral and listening skills," "to see how proficient students are in English – reading, writing, and speaking," "to identify the language developed by children, how much language they are using, how much is acquired," and "to assess students' understanding of English – can they follow directions." Some respondents stated that the current AZELLA KPT is simplified compared to the previous test and is "far more indicative of where children are."

Test Certification. The online training that resulted in the test certification received mixed comments from the interviewees. Most educators reported that the training was helpful, practical, and left them "knowing exactly what to do with a student." At the same time, some commented that they had not been initially informed that a 100% score on the training module quiz was required, and they had to go over the training more than once. Some test administrators were not certain about answers to some certification questions and thought that the training took too long. Several educators expressed their concerns about the late time of the training in the year and were hoping for an earlier certification process in the future.

Testing Process. When reflecting on the AZELLA KPT procedures, most participants reported following the directions that were presented in the online training. Many seemed to be supportive of the reduced length of the test, compared to the previous screening instrument, and noted that they were usually able to build rapport with a student on the way to the testing locations. Test locations varied across and within districts; some examples of testing locations listed by the interviewees were conference rooms, classrooms, libraries, empty offices, and – in some instances – hallways. Several test administrators voiced their preference for the stand-up flip-chart format of the testing booklet because it was easy to work with. Some also enjoyed giving students stickers at the beginning and the end of each test administration, although some pointed out that offering stickers at the beginning of the test was a distraction for some students.

Procedures for entering scores varied from district to district – with some test administrators entering scores themselves on the same day, others entering scores within the next several days, and still others delegating the data entry task to other staff members in their district. Some frustration was reported around electronic data entry due to system interruptions,

inability to identify and avoid duplicates efficiently (students' results across districts are not visible as entered, which is a challenge for large districts), and the possibility of errors in score entry. In response to some of these challenges, the interviewees recommended a process of uploading previously entered student scores or the possibility of using scannable bubble sheets. At the same time, several participants spoke favorably about the customer service support provided by the testing company. Also, a large number of interviewees were satisfied with the ability to obtain students' scores in an expedited manner. Still, some of them expressed support for more detailed and diagnostic reports. Although some educators noted that test reports were easy to read, others were not clear on the dual labeling of performance levels (e.g., basic-intermediate; incidentally, this level was reported to have a narrower range compared to the other performance levels). Many questioned the cut scores of the test and feared that some students may have been misidentified as proficient. One educator stated that "the previous test was too complex but this one is too easy"; another respondent thought that the AZELLA KPT was "a step in the right direction but we are not there yet."

Identification of Students in Need of ELD Programming. When asked to comment on the propensity of the AZELLA KPT to identify students in need of ELD programming, most participants seemed to believe that the test was helpful in identifying high achieving and low achieving students but they feared that some students in the middle may have been over-identified as initially fluent English proficient (IFEP). One participant noted, "The potential is there but the test might have missed some kids since the bar was a little low." Another interviewee pointed out, "We have been asking for a screener that does not underqualify students but it would also be helpful to have a screener that does not overqualify students." Interview findings point to educators' desire to have clear information about score cut-offs and how individual items are weighted.

Some participants were partially supportive of the accuracy of the testing results. One of them ascertained that the test identified students correctly in the areas of listening and speaking but raised concerns that the other domains, which may be more difficult for students, were not a focus of the test. Another participant believed that the AZELLA KPT was valid compared to the previous test that "qualified all students as ELLs." Both groups of test administrators – those who believed test results were accurate and those who questioned them – mentioned the upcoming Spring AZELLA test and were curious about how its results would compare to the initial AZELLA KPT scores.

Respondents were also asked to comment on any surprising aspects of the AZELLA KPT administration. The most frequently mentioned unexpected outcome was that some students were originally perceived as ELLs but they ended up passing the AZELLA KPT. Both the test administrators and some teachers they worked with seemed to share this insight. In a few cases, the respondents reported no unexpected outcomes.

Placement Decisions. In most cases, interview participants reported that no placement changes were made based on the AZELLA KPT results. They noted that the Primary Home

Language Other Than English (PHLOTE) survey had been used as the basis for making placement decisions. In one instance, classroom teachers were also consulted when students were identified for services. The primary reason for keeping all students in their original classrooms was reported to be the fact that the AZELLA KPT was administered when the school year was in progress, and students had already acclimated to their new classroom settings; therefore, it would have been undesirable, or even traumatic, for them to change classrooms once again. One respondent pointed out that no placement changes occurred in his or her school because all educators were SEI-endorsed and were able to tailor their instruction to ELLs' needs in every classroom. Several educators commented on the timing of the AZELLA KPT administration saying that an earlier administration would have allowed them to place students more appropriately. Some interviewees also noted that even though the AZELLA KPT did not lead to any placement adjustments in their districts, test results were still useful for teachers in making instructional decisions. This perspective on wanting to use the KPT to inform instructional planning was not in keeping with the original screening purpose of the assessment.

Insights About Students. To obtain a multi-dimensional picture of AZELLA KPT-related processes, some interview questions focused on students, namely on those types of students for whom the test appeared to be easy and those who seemed to experience difficulties during the test. The students perceived by administrators as high achieving were described as animated, interactive, articulate, talkative, and clear with their answers. Such students tended to produce detailed answers, draw on personal experiences, and tell personal stories. Three administrators mentioned that such students tried to answer some test questions even before the administrators were able to finish asking them. High performing students regarded the test as "fun and easy." Overall, such testing experiences were deemed to be pleasant, speedy, and easy to score. The question about whether such students produced enough language to demonstrate sufficient language proficiency, was mostly answered affirmatively. There seemed to be some relationship between the affirmative response to this question and perceptions of the AZELLA KPT as a screener rather than an English language proficiency test. Several other respondents thought that high achieving students should have generated more language for the assessors to be confident in their high achieving students' performance results.

Those students who struggled with the test were reported to be quiet, often silent, confused, shy, and hesitant. Test administrators indicated that they required more time before the test to build rapport with these students. Some students were able to offer answers, sometimes in another language; they also did not seem to understand directions. When students were completely silent, some administrators experienced a degree of frustration when having to deliver the entire test. For this group of students, the question about whether the amount of language students produced was reflective of their proficiency received mixed answers as well. Some educators thought that the language quantity was sufficient and others responded negatively to this question. Those who believed that students needed to generate more language pointed out that one-word answers were acceptable, and the test did not allow for much expressive language.

Test Items. One of the goals of the retrospective interview phase was to identify which test items were perceived by test administrators as easy and which test items were perceived as difficult. Most interviewees reported that the introductory self-description questions were easy for kindergarteners. Also, items entailing repeating, pointing to objects, drawing and coloring, as well as some basic questions on singular-plural nouns, were identified by respondents as easy test items.

The following were the more difficult items on the test, as perceived by the interviewees: some more complex singular-plural nouns, exact repeating of underlined words (according to several test administrators, those items were most difficult to hear and score adequately), retelling, sequencing of events, and identifying beginning sounds.

Administrators noted that in one section of the test, a part on initial word sounds, for all but the last item, the correct answer was B. Administrators often noted that students answered the last item in that section as B as well.

One administrator observed that it was positive that the test format changed frequently – this kept students attentive and engaged. Others had additional suggestions for improving test items:

- avoiding identical answer choice patterns
- including more items that measure students' command of academic language
- having more than one choice to color
- accounting for possible age-related speech impediments (e.g., due to missing teeth)

Although most participants reported that they followed the test administration scoring directions strictly, there were some exceptions. Findings from three interviews in one district indicated that in test items starting with "Tell me *about...*," the preposition was perceived as one prompting more than a one-word answer, and points for one-word responses were not awarded, which contradicted the certification instructions.

When asked whether the AZELLA KPT contained a sufficient number of items, most test administrators agreed that the length of the test was appropriate for the types of inferences made about students' proficiency levels as well as for the child's age. This finding was especially salient in the context of comparison to the previous AZELLA test which, according to some respondents, was much longer and less developmentally-appropriate. One respondent stated, "If you went longer you'd lose kids, shorter – you'd lose reliability and validity." Another test administrator noted that "much of the test was also watching the child," and the test format made this possible. Yet, some interviewees stated that the test cut scores needed to be adjusted. A few others were in favor of adding several more test items generating either more academic language or entailing pre-reading and pre-writing components. Still others believed that although the quantity of the items was appropriate, some of the items were repetitive, and the content of the test needed to be made more rigorous, again, with an emphasis on academic language used in the classroom.

Other Findings. The retrospective interviews generated the following additional findings:

- Some interviewees reiterated their preference for the new AZELLA KPT – they liked the individual testing format and considered the test to be developmentally-appropriate (because there was no writing section on the test). On the other end of this continuum, there were supporters of adding some pre-reading and pre-writing components to the test. A large number of educators pointed to the need for changing the test scale so that the test did not inadvertently identify students as English proficient when they were still in need of services.
- Many participants commented about the upcoming Spring AZELLA test – some were cautious about the early test administration and were wondering whether test results would be comparable to those produced by the AZELLA KPT. Others expressed concerns about implications of those results for (re)classifying students, school accountability ratings, SEI program funding, and staffing. Several educators also expressed caution about the speaker phone component of the spring test.
- Some interviewees identified a need for AZELLA-related professional development in-services for test administrators, teachers, instructional coaches, and other decision makers in the field. Some of the suggested topics were related to instructional strategies, alignment of instruction and assessment to Common Core and ELP standards, relation of the AZELLA KPT to other local and large scale assessments, and provision of more information about the AZELLA KPT to teachers, school administrators, and parents.
- Some test administrators were hopeful that future test administrations would start earlier in the school year so that students are identified and placed properly. Earlier timelines would also allow for better instructional planning and appropriate staffing.

Activity 3: Focus Groups

In order to get feedback from educators at school locations that were not included in other in-person data collection activities (e.g., retrospective interviews, case studies), asynchronous focus groups were held. The focus group also allowed project staff to gather emerging information from educators during the time frame in which the Spring AZELLA 2013 administration was ending and the assessment results were reported.

Focus Group Procedures

Asynchronous focus groups were conducted in an interactive online environment with a total of 10 test administrators and one teacher. Participants were recruited via a broad email invitation. Focus group participants were asked six consecutive questions over a period of four months (March-June, 2013). Participants were asked to both answer each question and comment on each other's responses. An one-day in-person meeting was held in the final month of the focus group period to give participants an opportunity to discuss Kindergarten assessment issues in-depth and reflect on some findings of the evaluation process.

The following are the six questions that were posed during the focus group process:

Question 1. Could you please tell us about how the administration of the AZELLA 3 went in your school/district? The AZELLA 3 followed the AZELLA placement test and was administered more recently, so we would like to learn a bit more about it first.

Question 2. Thinking back to the AZELLA Kindergarten Placement Test (KPT), did your school/district make any placement changes based on the test results? How is your school/district planning staffing for next year based on how things went this year?

Question 3. How would you like to see the KPT changed for future administrations? Please list up to 5 ways - these could include actual changes to the test, changes to the procedures for testing, and policies related to administration and placement.

Question 4. Are there any professional development opportunities around the AZELLA tests and related instruction and assessment that may be useful for you and/or your colleagues? If so, please list up to three topics of such professional development in-services.

Question 5. By now you have received your district's AZELLA 3 results. Tell us a little about your reaction to the results for Kindergarten. Are you surprised by your students' scores? Why or why not? Will there be any classification and placement decisions made based on these results?

Question 6. Since no changes to the KPT are anticipated this year, what are some resources that would be helpful for you in planning and administering the test this year?

Focus Group Findings

Question 1. The question about the administration of the AZELLA 3 was posted first because this assessment had been given recently, and participants were likely to share new perspectives. Most participants addressed issues of the AZELLA 3 administration related to the timing and scheduling of the test, including concerns about an early and short window for giving the test, challenges with the duration of the test, difficulties with giving a parallel test during the same window, and limitations of implementing the KPT and AZELLA 3 back-to-back for some incoming students. Although some administrators experienced technology setbacks, for others the testing went smoothly due to the comprehensive protocols for testing procedures prepared by the Arizona Department of Education and school districts. Some participants also expressed concerns about differences between KPT and AZELLA 3 results, which would have additional implications for school/district accountability.

Several participants addressed the administration of the speaking section of the test:

"The speaking portion was time-consuming since it could only be administered individually. However, I did see the consistency in the amount of time students received to formulate responses, regardless of their speed of response." (P10)

"Administering the speaking portion of the test was most challenging due to the requirement of a land line speaker phone as well as the time needed to administer the test. We also had one school that experienced difficulty when calling the phone number. The issue was resolved after phone calls to Pearson and our internal IT department." (P7)

"We didn't have any problems with the speaking test other than trying to find a place to do them and the number of tests for each staff member to administer." (P5)

Question 2. This question focused on the AZELLA KPT as well as subsequent placement decisions and staffing implications for next year associated with this assessment. Participants reported *that* most placement decisions were based on the PHLOTE survey results (and teacher observations). Some also noted that this year, higher numbers of students scored as "proficient" compared to the results in the previous years. Some educators faced challenges in communication with parents about discrepancies between KPT and AZELLA 3 results. Some also expressed concerns about staffing for next year based on AZELLA 3 results as well as concerns about using the KPT for new students arriving later in the year. One participant pointed out a need for information about the KPT that could be shared with teachers early next year.

The following insights in response to Question 2 were shared by some participants:

"The placement test went from one extreme of being very difficult on the old one for the beginning of the year to the other extreme of being very easy to pass it. ... There needed to be a happy medium for the Kinder screener." (P1)

"With the Common Core Standards now a requirement at the Kindergarten level, there is an increased cognitive demand for students in Kindergarten with regard to reading and writing. ... This placement test only measured basic communication skills yet students are required to perform academic tasks in English within the classroom setting." (P2)

"The most difficult thing ... is that, students are already placed in respective classrooms with respective teachers. By this time, usually after the first 10 days of school, they are accustomed to the learning structures, routines, and procedures of their teachers." (P10)

Question 3. When asked for suggested changes to the KPT, including changes to the test items or format, changes to the testing procedures, and policies related to administration and placement, some educators voiced a preference for a more academic version of the test that would incorporate rigorous pre-reading and pre-writing components. Several participants noted that the test scales, cut scores, or item weights needed to be adjusted so that it was not too easy for students to be identified as proficient when they still had language development needs. Some expressed a preference for a KPT that would have a greater alignment with the AZELLA 3. Some also found that handing out a sticker at the beginning of the test administration might

be distracting for some students. Most administrators reported that handing out a sticker at the end of the test proved very effective.

Several educators also pointed to the aspects of the KPT that worked well during the administration:

> "I think it was helpful to complete the test one-on-one. I think this helped the students focus on the test and not get distracted by other students in the room. As new students to school, this individualized approach to testing seemed appropriate and hopefully led to more accurate results." (P2)

> "I did like the beginning sound part of the test….this was very appropriate for the beginning of the year of Kinder." (P1)

> "We are delighted with our new ability to both enter results immediately and go in to enter SAIS ID #s that are missing at that time when they're received at a later time." (P4)

Question 4. This question generated topics for professional development in-services perceived to be useful by the participants. The list of the topics was used for the purposes of the Multi-Attribute Consensus Building activity, which is explained in detail later.

Question 5. When reacting to the AZELLA 3 results, study participants pointed out that fewer students passed the AZELLA 3 than the KPT. Although there appeared to be general consensus that the AZELLA 3 is more accurate in identifying ELLs than the KPT, classification and placement decisions varied (ELL, English language learner after reclassification – ELLAR, parent input, etc.). Of note, are the following quotes from the participants:

> "At first, I wasn't in agreement with a phone call giving a test, but after seeing the results, I think we need to be cautiously positive about this test." (P9)

> "I am happy to see that the results are more realistic and true to what I see in the students' English. For example, I saw one ELL who passed the Kinder Placement AZELLA and yet she spoke almost NO English (coming from Arabic). Now on the AZELLA she took in the Spring, she placed as an Intermediate which is pretty accurate by what I see (she is a very fast learner). The three who seemed IFEP passed the new AZELLA as Proficient, while the others who seemed to need a little more support tested as Intermediate. So I think the new AZELLA is much more accurate than the earlier Kinder Placement test." (P6)

Question 6. This question focused on the resources that would be helpful for educators in planning and administering the KPT next year. One participant noted that a refresher training on administering the test would be helpful in the upcoming academic year. Some participants also pointed out the benefits of the AZELLA 3 test, which helped identify some of the students missed by the KPT. Although some educators expressed their disappointment at the fact that no

changes were going to be made to the KPT this year, they understood that the process of revision is a lengthy one and requires a complex approach.

Multi-Attribute Consensus Building Results

The Multi-Attribute Consensus Building (MACB) process was employed to enable participants to discuss the professional development topics they generated online in response to Question 4, as well as during the in-person discussion, and attempt to reach consensus on which topics were most important to the participants. The MACB method is a quantitative approach for determining a group's opinion about the importance of each variable (in this case, each professional development topic) on a list. During the MACB process, all participants were asked to discuss and weight the importance of each professional development topic using the following MACB scale: 1-20 – very unimportant, 21-40 – unimportant, 41-60 – neither unimportant nor important, 61-80 – important, 81-100 – very important. Each participant was asked to assign a 100 weighting to at least one topic and compare the importance of the remaining topics to the top selected item(s). After all participants finished discussing and generating their weightings, each of them was asked to report his or her weightings for each professional development topic. The weightings were instantly entered into an MACB spreadsheet visible to all participants. The spreadsheet automatically calculated ranges, overall proportional weights, and importance averages for each professional development topic. These findings were subsequently discussed by the whole group with an emphasis on the topics that received more varied weightings. Participants were able to change any of their weightings at any point during the general discussion if they were convinced by others of a different level of importance of a given professional development topic.

Table 2 summarizes the results of the MACB process. According to the table, eight professional development topics were weighted as very important, and three topics were weighted as important. The following professional development topics received the highest average importance weightings: instructional strategies for Kindergarten ELLs; supporting proficient PHLOTEs; and CCSS alignment to ELP standards and early childhood methodologies.

Table 2. MACB Results: Importance of Professional Development Topics

Professional Development Topic	Range of Weightings	Proportional Weight	Average of Weightings
Instructional strategies for Kindergarten ELLs	85-100	0.20	98
Supporting proficient PHLOTEs	85-100	0.19	94
CCSS alignment to ELP standards and early childhood methodologies	80-100	0.18	91
Reading: differentiating in small groups via a balanced approach	75-100	0.18	89
Writing instructional strategies reflective of AZELLA 3 tasks	80-100	0.18	89
Implementing an ILLP	65-100	0.18	89
RTI model for ELLs in the context of the 4-hour block instruction	70-95	0.17	84
Training on the AZELLA 3	50-100	0.16	81
Communication with parents	60-100	0.16	80
Refresher training on the KPT	50-90	0.15	77
Documentation training (folder compliance)	50-90	0.15	74

Activity 4: Case Studies

In order to understand the effects of the KPT on a school district as a whole, NCEO staff conducted case studies that included classroom observations, interviews with Kindergarten teachers, and interviews with principals, other administrators, and coordinators of English language development services. NCEO staff spent approximately one week in each of the five districts that participated in the case studies.

Case Study Procedures

NCEO staff sought the assistance of the Arizona Department of Education in order to select districts for participation in the case studies. Selection criteria included a) having a sufficient number of English language learners enrolled in Kindergarten, and b) having schools in the district that were both high-performing and low-performing according to the state's accountability framework. A total of four districts and one public charter school were selected for the case studies. The charter school served as a site for piloting the instruments used in the

case studies. Although ADE helped identify the districts to participate in the case study, the identification of the districts was not retained by the state. District personnel who participated in the case studies were informed that ADE had assisted with identification, but that NCEO staff would not report findings in such a way that individual districts could be identified.

A team of two NCEO staff traveled to each case study location and spent approximately one week in the district. Two elementary schools were included in each site visit, and NCEO staff observed a minimum of two Kindergarten classrooms in each school. The duration of classroom observations was between 20 and 40 minutes. Instructional content included English language arts, math, health, and science. Classrooms were observed for use of academic language in English during instruction, based on an adapted protocol on academic language developed by Bailey, Butler, LaFramenta, and Ong (2004) (see Appendix D) as well as field notes focused on instructional activities. Individual students were not identified by name in the field notes.

Follow-up interviews were conducted with classroom teachers and principals. In addition, district English language development coordinators were interviewed. Other district personnel who participated in case study interviews included a district superintendent, data coordinators, an accountability coordinator, and KPT administrators. See Appendix D for observation and interview protocols used for the case studies. Field notes from the case studies were transcribed by project staff, using codes assigned to each case study location. NCEO staff conducted a thematic analysis of interview data, establishing themes that emerged across multiple case study locations. The results highlighted in this report are largely drawn from the case study interviews.

Background Information on the Schools Included in the Case Studies

A total of four districts and one charter school participated in the case studies. Locations included urban, suburban, and rural schools. Two districts were K-8 districts, and two districts served PK-12 students. The size of the districts ranged from fewer than 5,000 total students to more than 50,000 students. Similarly, the number of schools in the districts ranged from fewer than 10 to almost 100 schools. Spanish was reported as the most common home language for ELLs; still, as many as 31 languages were reported to be spoken by students in the case study locations. In the case study schools, Kindergarten students represented 15-30% of the school population. ELLs ranged from about 1% of students to more than 75% of students in a school. All schools in the case studies had over 75% of students receiving free or reduced lunch.

Models of English Language Development Observed

Across all schools, a total of 25 classrooms were observed, and included mainstream classrooms, structured English immersion (SEI) classrooms, pull-out ELD classrooms, and intervention hours conducted with mixed groups of ELL and non-ELL students. Some schools did not have enough students to have an SEI classroom in Kindergarten, so these students

received pull-out ELD through their Individual Language Learner Plan (ILLP). Figure 4 shows the breakdown of the classrooms observed.

Figure 4. Types of Classrooms Observed (n=25)
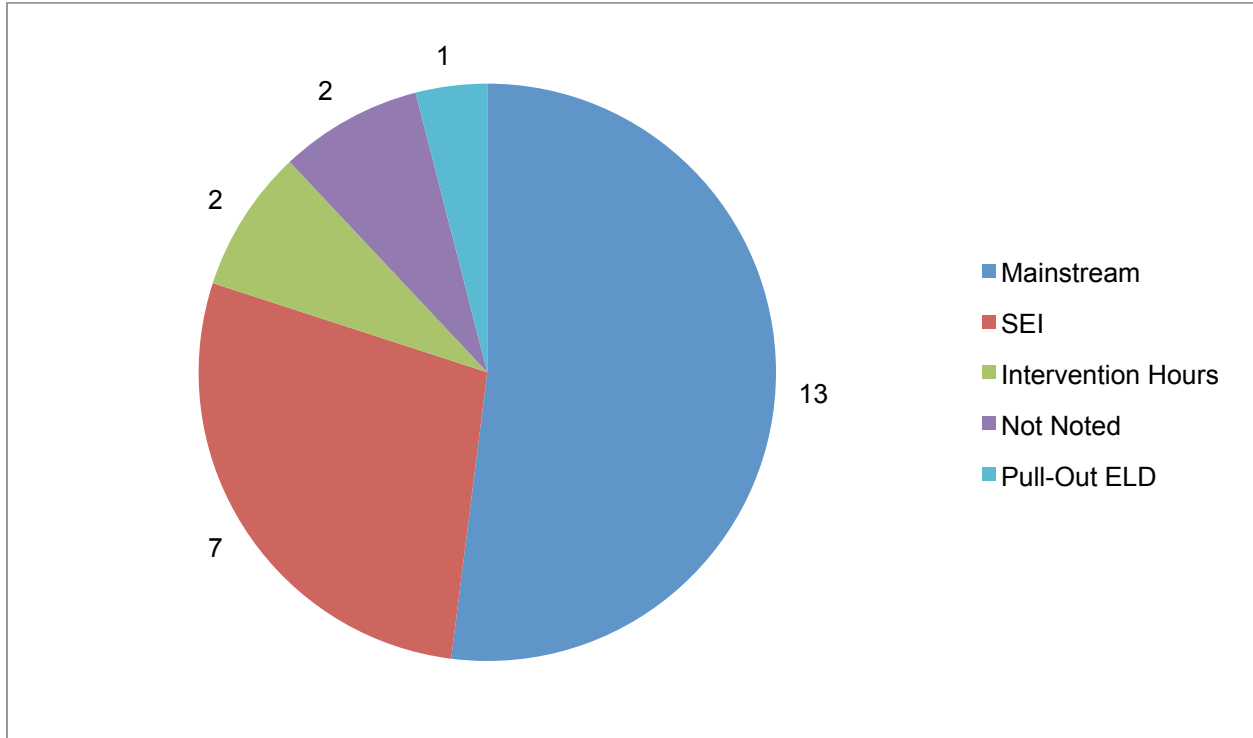
Generally, classrooms that were observed had 20 or fewer students. Mainstream classrooms had, on average, 20.6 students. SEI classrooms had an average of 18.7 students. Figure 5 provides additional information on the average number of students in the classrooms that were observed. Only one section of pullout ELD services was observed.

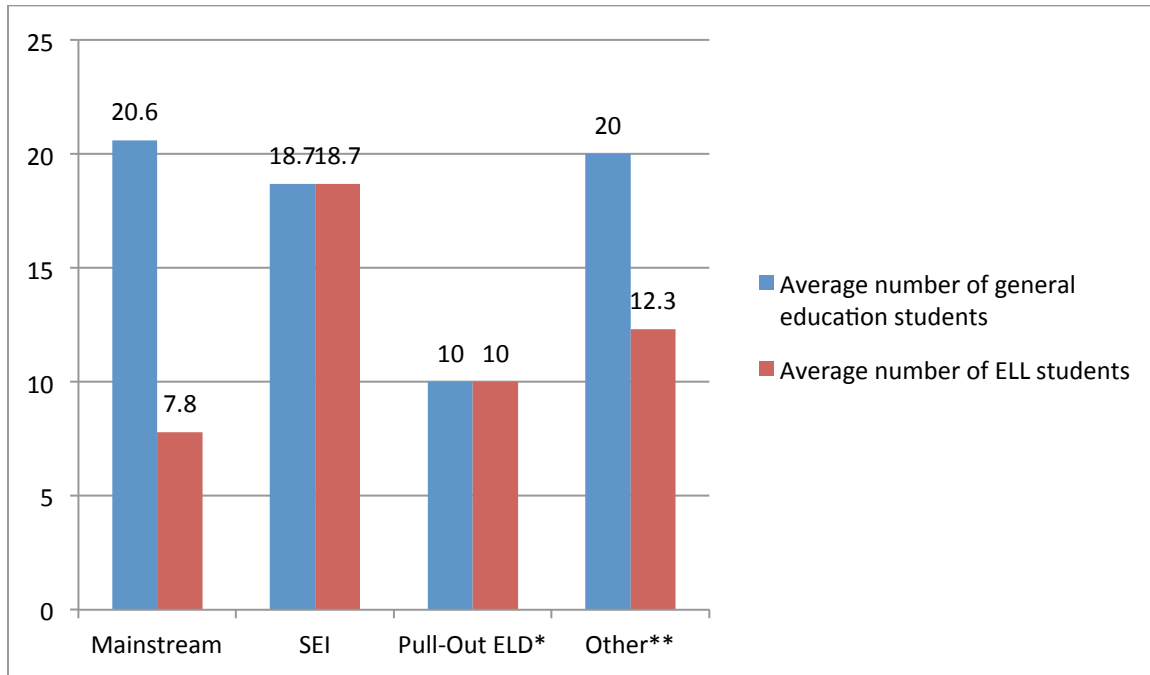**Figure 5. Average Number of Students in Class by Type**
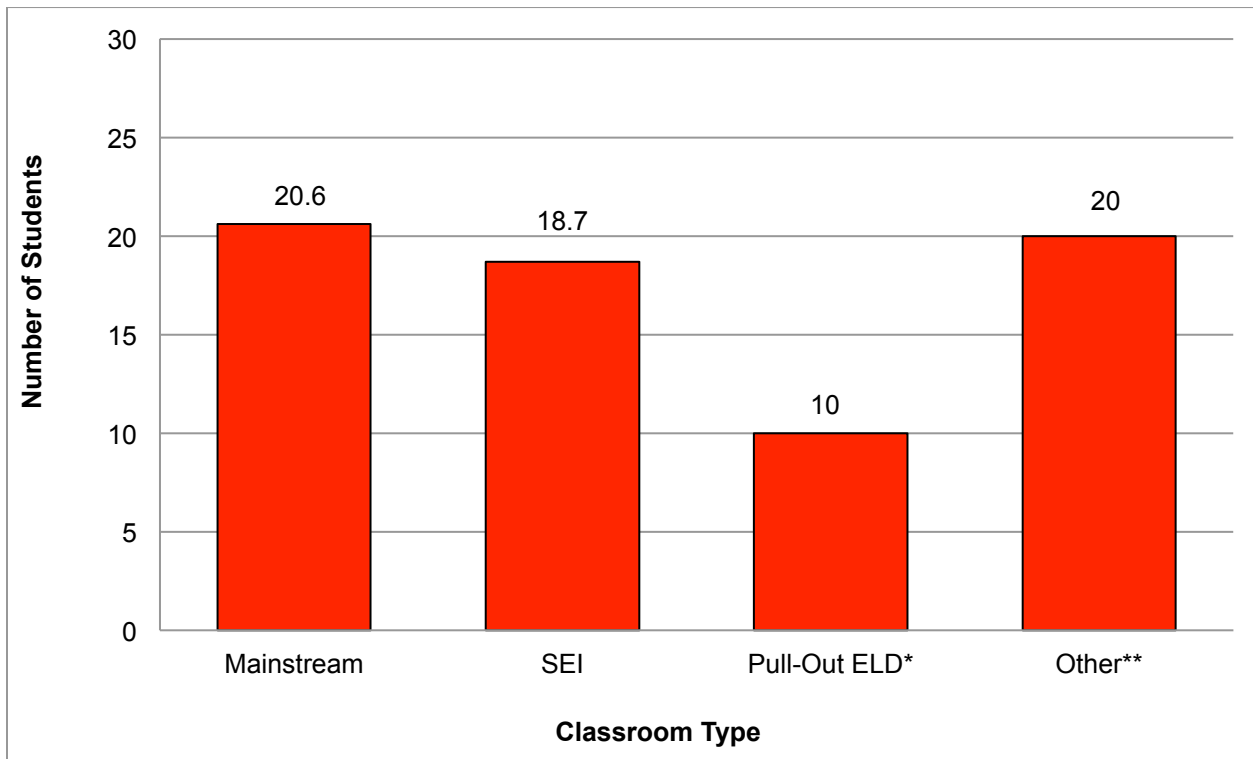
*Data are from one classroom only
**Category includes two classrooms providing intervention hours, and two classrooms in which type of class was not noted

In classrooms that had both mainstream students and ELLs, there was, on average, a higher number of mainstream students than ELLs in each class. Figure 6 illustrates the average number of mainstream students and ELLs in each class type. Note that the "Other" category primarily consisted of intervention hours that had a number of small group activities going on within the same classroom. These groups typically included both ELLs and mainstream students.

Figure 6. Average Total Number of General Education Students and ELL Students in the Classroom
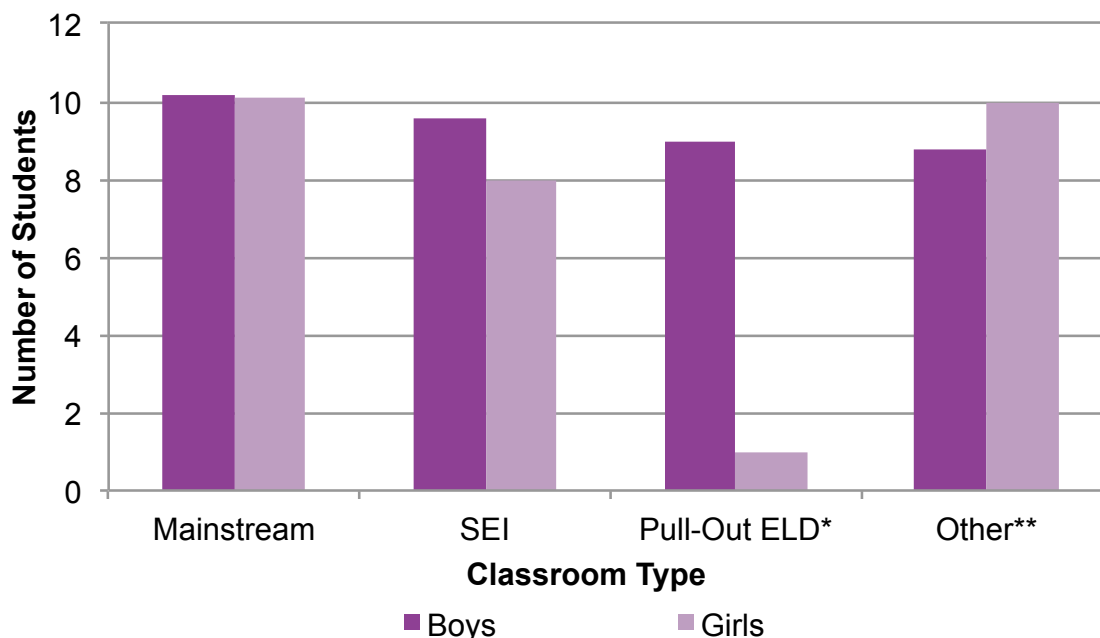

*Data are from one classroom only
**Category includes two classrooms providing intervention hours, and two classrooms in which type of class was not noted

Most classrooms serving ELLs had even representation of boys and girls, with the exception of the one pullout ELD class observed. Figure 7 shows the representation of gender across the classrooms observed.

Figure 7. Average Number of Students in Classrooms by Gender

*Data is from one classroom only
**Category includes two classrooms providing intervention hours, and two classrooms in which type of class was not noted
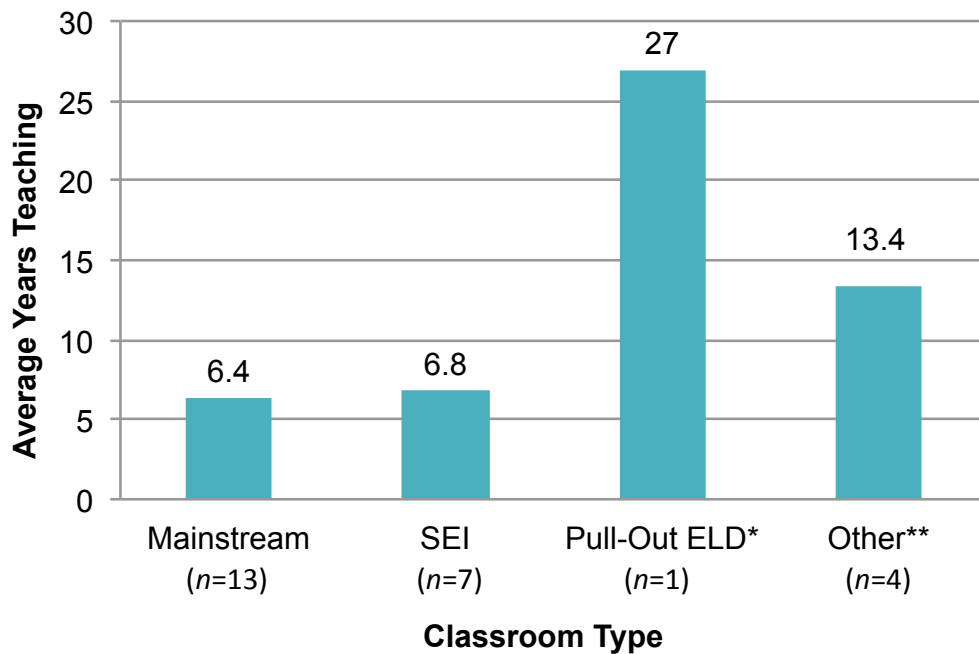
Classrooms lessons observed included English language arts, math, science, and health. Specific topics of lessons included writing stories based on pictures, blending words, story prediction, addition using a number line, brushing teeth, and oceans. In the intervention hours observed, students participated in a variety of small group English language arts instruction.

Background Information About Teachers Interviewed. After NCEO staff conducted classroom observations, participating teachers were interviewed using semi-structured interview protocols. Teachers were asked to describe the lesson that had been observed. Additional questions focused on teachers' understanding of academic language, their perceptions of the effectiveness of SEI certification, and their expert judgment related to the performance of ELLs, including those students who tested as Initial Fluent English Proficient (IFEP).

A total of 25 teachers participated in the case study interviews. Of these, 23 were female and 2 were male. One educator who participated in the interviews was a long-term substitute teacher working on teacher certification; the remaining educators were certified teachers. With the exception of the long-term substitute teacher, all educators interviewed had either completed the SEI certification or were in the process of finishing the certification. Most of the educators interviewed also had an early childhood certificate.

Figure 8 shows the average years of educator experience. Teachers ranged from having roughly one-half year of experience to 29 years of experience on the job. Those teachers who taught in SEI classrooms had more years of experience, on average, than mainstream teachers. This was supported by information collected in principal interviews, with principals noting that they tended to assign their more experienced educators to SEI classrooms.

Figure 8. Average Number of Years of Educator Experience

Average Years Teaching (y-axis: 0, 5, 10, 15, 20, 25, 30)

Classroom Type	Average Years Teaching
Mainstream (n=13)	6.4
SEI (n=7)	6.8
Pull-Out ELD* (n=1)	27
Other** (n=4)	13.4

Classroom Type

**Category includes two classrooms providing intervention hours, and two classrooms in which type of class was not noted

Figure 9 shows percentages of ELLs by classroom type. As shown in the figure, mainstream classrooms included 38% ELLs, SEI and pull-out ELD classrooms included 100% ELLs, and other classrooms (two classrooms offering intervention hours and two classrooms with unspecified instruction) included 61.5% ELLs.

Figure 9. Percentage of ELLs in the Classroom by Type
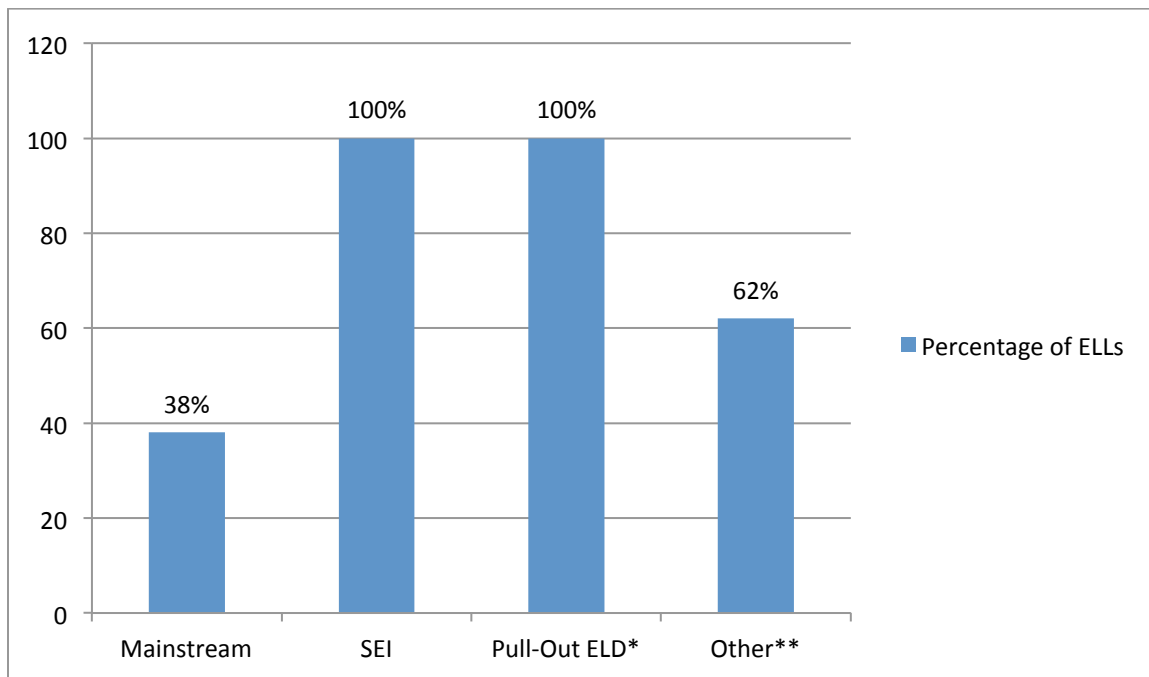
*Data are from one classroom only
**Category includes two classrooms providing intervention hours, and two classrooms in which type of class was not noted

Understanding of Academic Language. The Arizona Department of Education provides a definition of academic language on its website:

> ACADEMIC LANGUAGE: broadly defined, includes the language students need to meaningfully engage with academic content within the academic context. Academic language includes the words, grammatical structures, and discourse markers needed in, for example, describing, sequencing, summarizing, and evaluating—these are language demands (skills, knowledge) that facilitate student access to and engagement with grade-level academic content"
> (http://www.azed.gov/wpcontent/uploads/PDF/ELPGlossary.pdf).

This definition is the one that informed the development of English proficiency standards and the AZELLA assessment.

Teachers tended to define academic language in a variety of ways that were somewhat different from the state definition. Some referred generally to the idea of register (e.g., language not used on the playground). Others referred to the setting where language is used (i.e., the language of the classroom) but did not give details of specific aspects of language on

which to focus. Still others did list specific aspects that they believed defined academic language; these usually focused on vocabulary. A few teachers described academic language in terms of language modalities such as listening, speaking, or reading. Very few teachers mentioned writing as a component of academic language. One teacher added a behavioral component noting that academic language included knowing how to behave in school. A few teachers added specific knowledge they believed the students needed to have in English in order to have academic language. These skills included math skills, knowing letters, knowing numbers up to 20, knowing consonant-vowel-consonant patterns, knowing phonograms in reading and having the ability to decode words. One teacher implied that academic language included having conceptual background knowledge in order to understand the subject matter of the classroom.

By far the most common views were that academic language was the language of the classroom, that it did not include social language (i.e., Basic Interpersonal Communication Skills), that it involved different vocabulary from that used on the playground or in everyday life, and that reading and listening comprehension were key aspects of academic language. Teachers generally did not refer to language functions such as describing, sequencing, and functioning that are included in the state definition of academic language.

Perceptions of SEI Certification. Educators were asked about their perceptions of the SEI certification and whether they believed it was sufficient in preparing them to teach ELLs and in SEI classrooms. A total of 21 teachers provided feedback on the SEI training. Four teachers were completely positive in their responses, five thought the training was adequate or listed aspects of the training that had helped them but did not directly state that they liked or disliked the training. Seven teachers had mixed feelings about the training. Generally these comments were of two types. First, several teachers indicated that the training was a good start, but it did not go far enough in either the depth of the content or the duration. Some teachers asked for additional training and one believed that ongoing training in small chunks would be preferable to a one-time training. Another teacher asked for a corresponding practicum in SEI teaching after the training. Second, some teachers thought that the training was generally adequate to start with, but that they needed to go back to their classrooms and gain hands-on experience in order to fully comprehend what they had learned. Three teachers responded with negative comments about the SEI training. Two believed that the training they received was simply not enough to meet their needs. One indicated that there was too much information provided and that the information was too hard to digest. In general, teachers would like more instructional strategies to use in SEI classrooms.

Comments From Teachers on the KPT Including Performance of Initial Fluent English Proficient (IFEP) Students. Prior to the classroom observations, teachers provided NCEO staff with classroom seating charts. Students were not identified by name, but their ELL status (e.g., current ELL or IFEP) was noted. Students were designated as IFEP if they were initially identified to take the KPT in the fall but received a score on the screener that indicated they should not be designated as ELLs. Educators were asked about the performance of IFEP

students in their classrooms. Educators reported that the majority of IFEP students were performing satisfactorily in mainstream classrooms. Educators reported that they were paying attention to these students' performance on other measures, such as Dynamic Indicators of Basic Early Literacy (DIBELS), a short formative measure of early literacy. They also mentioned that they were interested in seeing these students' Spring AZELLA scores so that they could consider placement options for first grade. It should be noted that educators also mentioned that they believed some shy students were inappropriately classified as ELLs according to the KPT. However, educators did not report examples of harm that students experienced as a result of having been classified as ELL or as IFEP.

Background Information From Principals Interviewed. A total of nine principals (including the charter school administrator) were interviewed as part of the case studies. Interviews were conducted in the case study schools, using a semi-structured interview format. The majority of principals was male (n=6); 3 female principals interviewed. Principals ranged in experience, from 7 years to being in their first year of principalship.

Principals generally reported that they found out about the KPT administration from the ELD coordinator for the district. Principals did not participate in the KPT administration and typically reported a limited understanding of the KPT itself.

Principals also expressed concerns about the administration of the 2013 AZELLA 3. Specific concerns focused on the timing of the administration in January and February, rather than later in the school year. Principals described logistical concerns they had encountered with the phone testing requirement for the speaking test. In particular, they commented that many schools were not equipped with the appropriate type of speaker phone for students to use in completing the speaking portion of the test. This concern was not limited to the Kindergarten administration.

Communicating Results to Parents. Principals were asked about how they communicated the KPT results to parents. Most principals described sending a letter home with students to share with parents. Other principals noted, and teachers confirmed, that KPT results were shared with parents in fall conferences. Although some principals observed that a few parents were surprised by their child's results, most principals did not note unusual parent reactions to KPT results.

In one district, a concern was raised about retesting students who were IFEP. Specifically, the parent information provided by the state indicated that the student would not be tested for English language proficiency again. However, in the spring of 2013, all Kindergarten IFEP students also took the Spring AZELLA 3. In this district, the decision was made not to include the information about no additional language testing, and this information was also communicated to ADE.

Staffing Issues Encountered. In two of the case study districts, principals reported making placement changes based on the results of the KPT administration. In one district, one

Kindergarten teacher changed schools due to shifts in enrollment. In the same district, a long-term SEI teacher was shifted to teaching a general education classroom that included several ELLs on Individual Language Learning Plans (ILLP). Principals reported concerns about future staffing dependent on student performance on the 2013 Spring AZELLA 3 administration.

IFEP Students. Principals expressed concern about the performance of all students, including IFEP students, on the Spring AZELLA 3 administration. Principals noted that they received bonus points in the state's accountability model, for reclassification of ELLs. Many principals expressed concern that they would not receive those bonus points, which would potentially influence whether the school would maintain its rating or receive a lower rating. They commented frequently about public perception of the school's accountability rating.

Principals also expressed uncertainty about the performance of IFEP students on the Spring AZELLA 3 administration. Principals were unsure about placement decisions for IFEP students who would not score proficient on the Spring AZELLA 3. Most principals reported that these placement decisions would be made on an individual basis.

The concerns raised by principals in the case study schools prompted many of the questions in the principal survey. See Activity 5 for more information on how principals across the state addressed the concerns raised by the changes in ELL populations related to the KPT.

Activity 5: Principal Survey

During the 2012-2013 retrospective interviews and case study data gathering, NCEO staff learned that Arizona principals were concerned about the effect of the Arizona KPT and the Spring AZELLA 3 administration on their school accountability scores. NCEO conducted a survey of Arizona principals in the fall of 2013 to learn more broadly what the concerns of Arizona principals were about the Kindergarten language proficiency assessments.

Principal Survey Procedures

The Arizona Principal Survey was a 25-item electronically administered survey measuring Kindergarten principals' opinions of the administration of the Arizona KPT and subsequent AZELLA 3 administration during the 2012-2013 school year. The complete survey is included in Appendix E. The Arizona Department of Education provided NCEO with a list that included all principals at public and charter schools in Arizona. A total of 1,033 Arizona Kindergarten principals were sent the electronic survey on September 3, 2013; two reminder e-mails were sent, one on September 10, 2013, and another on September 26, 2013. Of all recipients, 181 principals responded to the survey generating a response rate of 18%.

The number of participants responding to each item varied because not all items were applicable to every participant; in addition, some participants did not answer every question.

Figures that depict results include only those participants who responded to that particular item, thus N sizes vary across the figures.

Demographic Data

The introductory section of the survey included some demographic questions intended to generate a better understanding of the respondents and their schools. Figure 10 shows the percentages of the principals' schools that enrolled ELLs. According to the figure, 93% of respondents reported that their schools served ELLs. The remaining 7% responded that their schools did not serve ELLs; these respondents were excluded from the results of the remaining survey questions.

Figure 10. Percentage of Schools that Serve ELLs

- Serves ELLs (N=163)
- Does not serve ELLs (N=13)
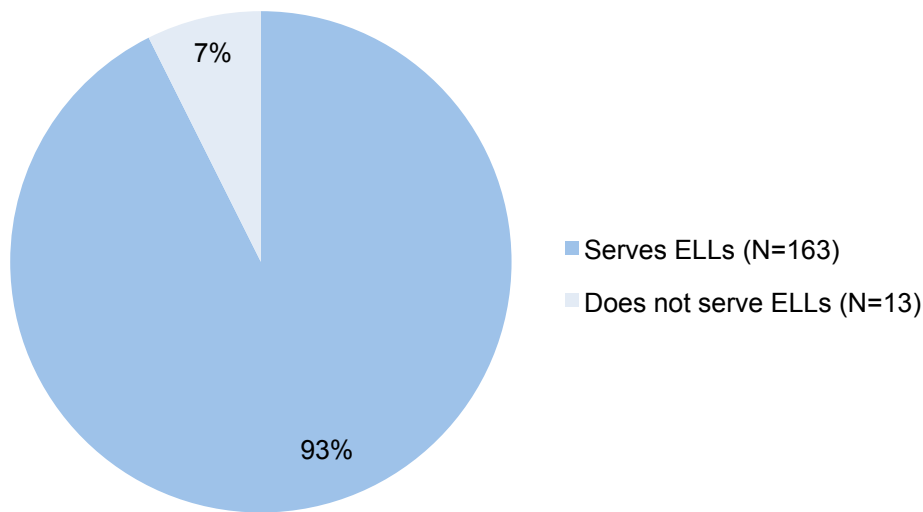
Another demographic question of the survey focused on the types of schools that survey respondents represented. Figure 11 shows these data. The majority of schools represented in this sample (66%) were reported as district schools. Another 33% were reported to be charter schools (with one school reported in the *Other* category as a charter school under a district jurisdiction).

Figure 11. Percentage of Types of Schools in Which Participants Work

- District (N=111)
- Charter (N=55)
- Other (N=1)
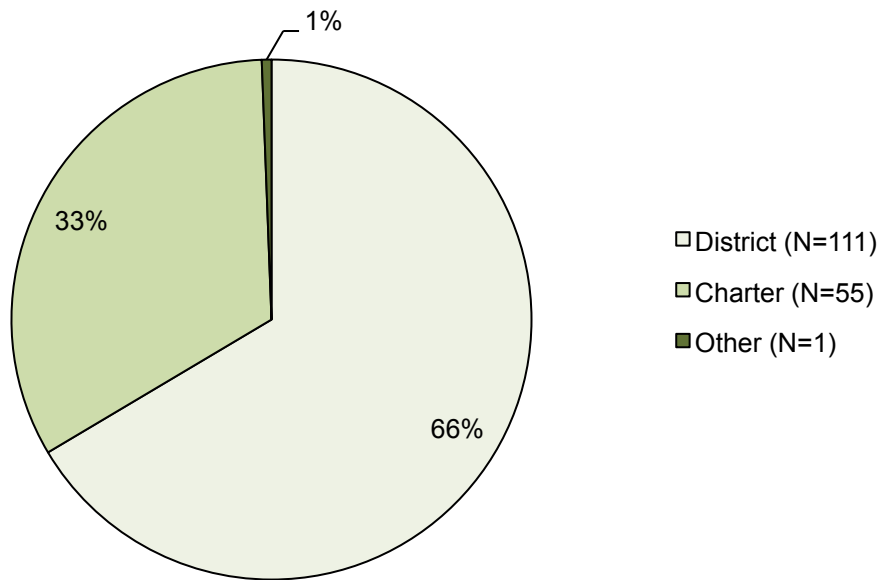
Survey Findings

This section highlights findings generated by the survey. Table 3 shows reasons principals provided for including IFEP students (on the 2012 KPT) in mainstream classrooms in 2013-2014 if they had been identified as not proficient on the AZELLA 3 in Spring 2013. The top reason that was cited by 50% of respondents was the fact that such students appeared to have the skills needed for instruction in mainstream classrooms. A smaller group of respondents (11%) stated that test results were reported to schools after classrooms were assigned, and reassigning students would be disruptive, while 1% pointed to the impracticality of meeting with so many parents to change students' placements. The remaining 38% comprised the *Other* category which, as well as the *Other* responses noted below, will be addressed in a separate principal survey report.

Table 3. Reasons for Including Students in Mainstream Classrooms

Decision Reasons	Number of Participants	Percent of Participants
Students have the skills needed for inclusion in mainstream classrooms	80	50%
Scores arrived after classrooms were assigned and moving students would be disruptive	17	11%
Meeting with parents was not practical	2	1%
Other	61	38%
Total	**160**	**100%**

School principals were also asked some questions about staffing implications of the results from the KPT and AZELLA assessments. One of the questions addressed the availability of ELD teachers – both SEI and Bilingual – for the 2013-2014 school year. As shown in Figure 12, 65% of respondents noted that they had sufficient numbers of ELD teachers for the upcoming year, while 17% required additional staff members. The *Other* category responses totaled 18%. It should be noted that at the time of the survey, the 2013-2014 school year had already begun, and so, presumably, most staffing concerns had already been addressed for that particular school year.

Figure 12. Availability of ELD Teachers for the 2013-2014 School Year
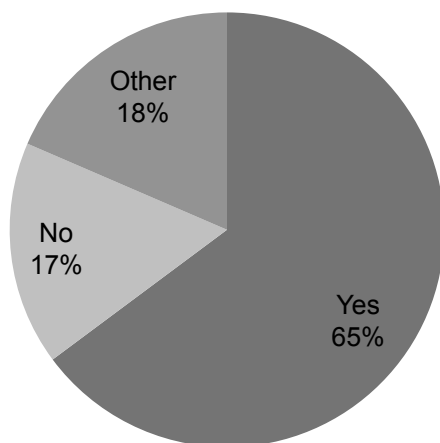
Another question focused on staffing asked whether survey participants had to get general education teachers to teach ELD (SEI or Bilingual) classrooms in the 2013-2014 academic year. Twenty-five percent of school principals answered this question affirmatively, and 65% of school principals responded negatively. The remaining 10% of responses were in the *Other* category. Figure 13 shows this information.

Figure 13. General Education Teachers in ELD Classrooms
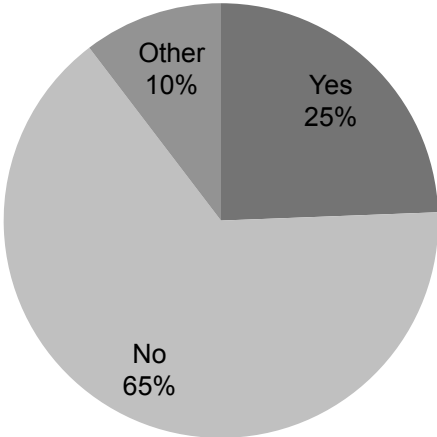

Figure 14 shows the results generated by the question about Title III funding for all students in ELD (SEI or Bilingual) classrooms for 2013-2014. Thirty-six percent of responding school principals reported that they received such funding for all students, 46% of school principals responded negatively, and the answers from 18% of school principals were placed in the *Other* category.

Figure 14. Title III Funding for All Students in ELD Classrooms
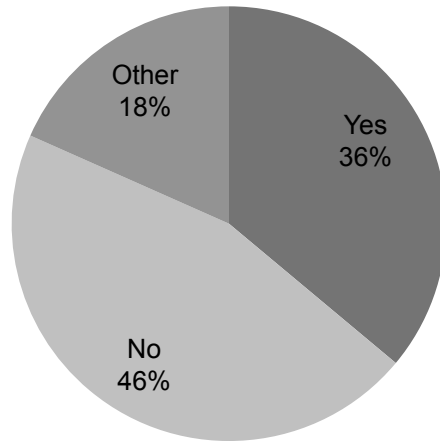
The answers to the question about whether AZELLA 3 results were being communicated to parents distributed the following way: 95% responded that test results were being communicated, 1% responded that test results were not being communicated, and 4% answered in the *Other* category. Figure 15 summarizes these results.

Figure 15. Communicating the 2013 AZELLA 3 Results to Parents
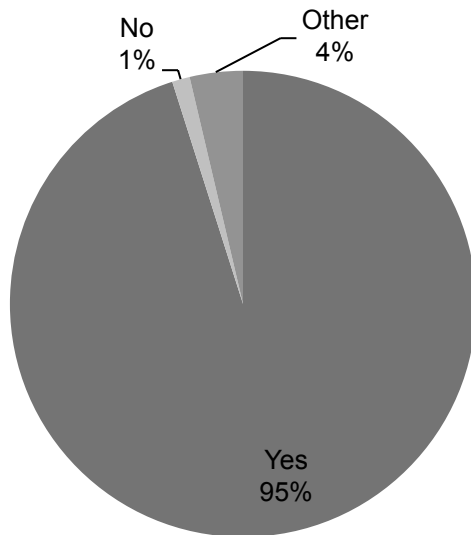
The school principal survey also included a question about whether parents had made inquiries or expressed concerns about the results of the 2013 AZELLA 3. Figure 16 shows that 26% of

respondents answered this question affirmatively, 71% of respondents answered that parents had no questions or concerns, and 3% answered in the *Other* category.

Figure 16. Parents' Questions or Concerns About the 2013 AZELLA 3
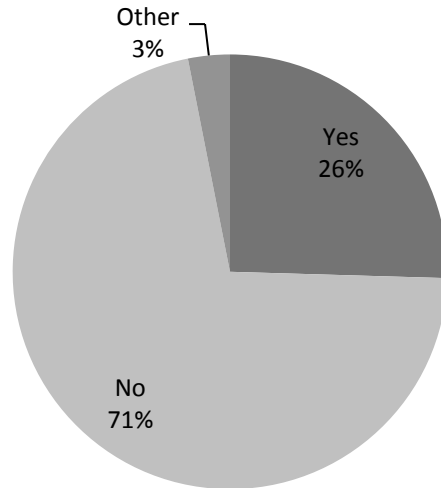
Activity 6: Data Analysis

A final activity conducted by NCEO was data analysis of the performance of students on the KPT and the Spring 2013 AZELLA 3 administration.

Data Analysis Procedures

NCEO initially obtained data from ADE for the schools included in the case study; after conducting a preliminary analysis with these case study schools, NCEO project staff requested data for the state Kindergarten population. SPSS was used to run descriptive analyses to learn more about the performance patterns among different groups of students taking the KPT and 2013 AZELLA. Findings here are for the state Kindergarten population. The analyses presented here do not include a breakdown of the performance of ELLs on the subtests of the AZELLA.

The three guiding questions for these analyses were:

1. What was the Spring 2013 AZELLA performance of students who were proficient on the KPT in the Fall of 2012?
2. What was the Spring 2013 AZELLA performance of students who were not proficient on the KPT in the Fall of 2012?
3. How would an increase in the KPT proficiency cut score potentially change the percentage of students scoring proficient on the AZELLA?

We present summary findings for each question. For the first two questions we used the AZELLA overall proficiency level rather than the proficiency level for a specific domain (reading, writing, listening, and speaking). To be considered English proficient for the overall proficiency level, a student must receive a rating of proficient in reading, writing, and the total combined AZELLA score across domains. Overall proficiency levels are not reported as scale scores.

Students who had both KPT and AZELLA scores. Table 4 shows that 20,398 students had scores on both the KPT, administered in Fall 2012, and the Spring 2013 AZELLA assessment. More than one third of these students (38%, n=7,825) were below proficient on both assessments. Approximately 32% (n=6,552) of them were at or above proficient on the KPT, but below proficient on the AZELLA, and 23% of them were at or above proficient on both assessments. Approximately 6% (n=1,272) of these students who were below proficient on the KPT, were proficient on the AZELLA.

Table 4. A Comparison of Student Proficiency/Non-Proficiency on the KPT and Spring AZELLA Assessments

		KPT Performance (tested n=21,054)					
		At or Above Proficient		Below Proficient		Total with KPT Score	
		Number	Percent of Total[a]	Number	Percent of Total[a]	Number	Percent[a]
AZELLA 3 Performance[b]	At or Above Proficient	4,749	23.3%	1,272	6.2%	6,021	29.5%
	Below Proficient	6,552	32.1%	7,825	38.4%	14,377	70.5%
	Total with AZELLA 3 score	11,301	55.4%	9,097	44.6%	20,398	100.0%

Note. [a] The denominator of the percentages in Table 4 is the total number of students with both KPT and AZELLA 3 Overall Scores (n= 20,398). [b]A total of 656 students took the KPT but did not have an AZELLA 3 score.

AZELLA Performance of Students Who Were Proficient on the KPT. Of the 11,301 students were proficient on the KPT and who took the Spring 2013 AZELLA, 4,749 of them (roughly 42%) were "at or above proficient" on both the AZELLA and the KPT. In contrast, 6,552 of these 11,301 students (roughly 58 %) were not proficient on the AZELLA, but had been proficient on the KPT.

Additional analyses, not shown here, indicated that of the students who were proficient on both assessments (the KPT and the AZELLA), listening and speaking were the two areas where students were more likely to score lower. This most likely is due to the fact that receiving an overall proficient level score on the AZELLA requires proficiency in the domains of reading and writing, as well as an overall proficient composite score, but not listening and speaking. Students who were proficient on the KPT and were not proficient on the Spring AZELLA were more likely to have lower performance in the domains of reading and writing than their peers who were proficient on both tests.

Performance of Students Who Were Non-Proficient on the KPT. Of the 9,097 students who were not proficient on the Fall KPT, roughly 86% of these students (n=7,825) were not proficient on the Spring AZELLA, and approximately 14% (n=1,272) were at or above proficient on the Spring AZELLA. Additional analyses, not shown here, indicated that 48 students in this group were most likely to have lower scores in listening and speaking compared to reading and writing.

Table 5 provides an analysis of how many students who tested during the 2012-2013 school year would have been proficient on the AZELLA if the cut score of the KPT was raised from 32 out of 42 points (the current cut score) to a higher number of points. The row showing data for the current cut score in the table is highlighted in grey. The numbers in parentheses in the percentage column indicate the changes in proficiency status on the AZELLA based on changes to the KPT cut scores. The analysis was not conducted using a potential cut score of 41 or 42 points because that would indicate that the student would need either a perfect score or could only miss one point to be considered proficient on the KPT.

For each potential increase in the KPT cut score (in the far right column), the percentage of students "at or above proficient" on both the KPT and the Spring AZELLA increases by approximately 1.5% to 4.5%. The increase is fewer than 2% when the cut score increases from 32 to 33, from 33 to 34, or from 34 to 35. The size of the increase jumps to 2.2% at a cut score of 36 and continues increasing by larger percentages from that point on.

This table suggests that raising the cut score by 3 to 4 points on the KPT could have an effect on the number of students who score proficient on the Spring AZELLA. Figure 17 graphically presents similar information to that shown in Table 5. The figure shows students who were proficient on the KPT and not proficient on the AZELLA in bar form, with the x-axis representing the cut score.

Table 5. The Relationship Between Increasing KPT Cut Scores and Changes in the AZELLA Overall Proficiency Level

| | | KPT Proficiency Status | | | | | | KPT Cut Score |
| | | Below Proficient | | At or Above Proficient | | Total | | |
		Number	Percent	Number	Percent	Number	Percent	
	Below Proficient	13,396	75.4%	981	37.2%	14,377	70.5%	
	At or above Proficient	4,367	24.6%	1,654	**62.8% (+4.5%)**	6,021	29.5%	40
	Total	17,763	100.0%	2,635	100.0%	20,398	100.0%	
	Below Proficient	12,721	77.4%	1,656	41.7%	14,377	70.5%	
	At or above Proficient	3,709	22.6%	2,312	**58.3% (+3.8%)**	6,021	29.5%	39
	Total	16,430	100.0%	3,968	100.0%	20,398	100.0%	
	Below Proficient	11,960	79.3%	2,417	45.5%	14,377	70.5%	
	At or above Proficient	3,129	20.7%	2,892	**54.5% (+2.9%)**	6,021	29.5%	38
	Total	15,089	100.0%	5,309	100.0%	20,398	100.0%	
	Below Proficient	11,191	81.0%	3,186	48.4%	14,377	70.5%	
	At or above Proficient	2,619	19.0%	3,402	**51.6% (+2.6%)**	6,021	29.5%	37
	Total	13,810	100.0%	6,588	100.0%	20,398	100.0%	
AZELLA Overall Proficiency Status	Below Proficient	10,433	82.3	3,944	51.0%	14,377	70.5%	
	At or above Proficient	2,239	17.7%	3,782	**49.0% (+2.2%)**	6,021	29.5%	36
	Total	12,672	100.0%	7,726	100.0%	20,398	100.0%	
	Below Proficient	9,709	83.5%	4,668	53.2%	14,377	70.5%	
	At or above Proficient	1,917	16.5%	4,104	**46.8% (+1.7%)**	6,021	29.5%	35
	Total	11,626	100.0%	8,772	100.0%	20,398	100.0%	
	Below Proficient	9,066	84.6%	5,311	54.9%	14,377	70.5%	
	At or above Proficient	1,656	15.4%	4,365	**45.1% (+1.7%)**	6,021	29.5%	34
	Total	10,722	100.0%	9,676	100.0%	20,398	100.0%	
	Below Proficient	8,433	85.3%	5,944	56.6%	14,377	70.5%	
	At or above Proficient	1,456	14.7%	4,565	**43.4% (+1.5%)**	6,021	29.5%	33
	Total	9,889	100.0%	10,509	100.0%	20,398	100.0%	
	Below Proficient	7,825	86.0%	6,552	58.0%	14,377	70.5%	
	At or above Proficient	1,272	14.0%	4,749	42.0%	6,021	29.5%	32
	Total	9,097	100.0%	11,301	100.0%	20,398	100.0%	

Note: The parentheses ( ) under the percent in the "at or above proficient" column and row indicate the percentage increase in students achieving proficiency for that particular cut score, compared to the percentage shown at the previous cut score.

Figure 17. Percentage of Students Proficient (PF) and Non-Proficient (NP) on AZELLA by KPT Cut Score
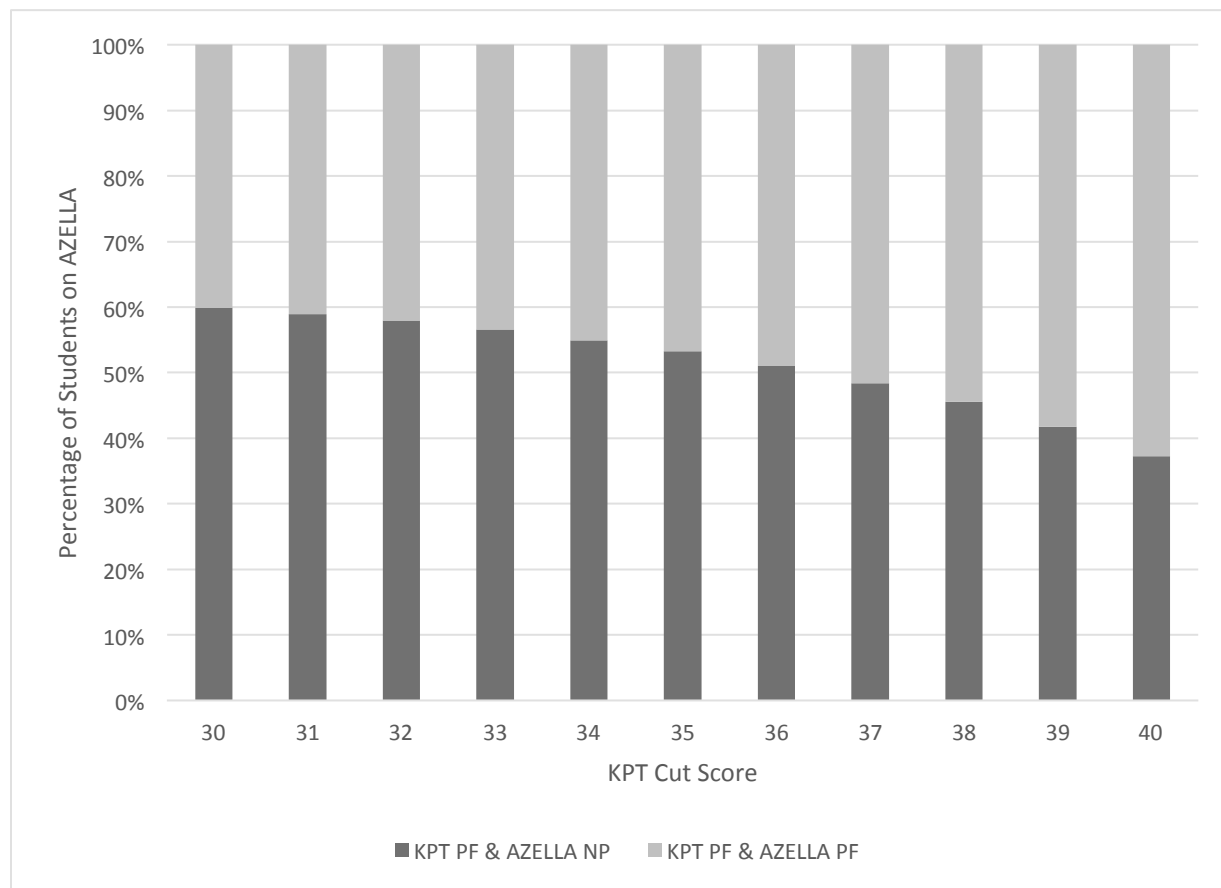
Summary of Data Findings. The analysis NCEO conducted indicated that a majority of students who passed the KPT did not receive an overall score of proficient on the Spring AZELLA. A small number of students who did not pass the initial KPT did receive an overall score of proficient on the Spring AZELLA. A cut score analysis suggested that a higher cut score of at least 35 or 36 points might be beneficial if the state prefers to reduce the numbers of students who are initially identified as fluent English proficient by one test and later the same year are identified as an English language learner by another test.

Recommendations

Based on the findings from the five activities undertaken by NCEO, a number of recommendations from the NCEO project team are presented:

Increased Communication With Districts About the KPT. In general, interviews indicated that school personnel were not clear about the purpose of the test nor about the procedures and timelines associated with the test. NCEO recommends that the Arizona Department of Education increase its communication related to the KPT with districts. We suggest that in addition to providing more communication, other stakeholders, such as principals and superintendents, be given more information about the KPT and its intended purpose.

For assessment coordinators and other district personnel administering the KPT, additional information about test procedures may be beneficial. Best practices for assessment administration for young children should be established to encourage test administrators to use child-friendly furniture and to test in quiet locations. Furthermore, the state may want to encourage test administrators to establish clear procedures for entering scores.

Timing of the Administration of the KPT. In the first year of the KPT test administration, there was not much time for test administrators to complete the training in time to give the test to students before the school year started. The testing window needs to be prior to the school year for school officials to use the test for placement decisions.

Adjustments to the Test Itself. Questions and concerns about several items were sometimes the result of limited understanding of the test blueprint. Still, other findings may indicate that there is a need to re-evaluate test items. For example, in one section of the test, nearly all of the answers are B. We recommend that this section be reconsidered.

Reconsideration of the KPT Proficiency Cut Scores. From the observations and interviews, our recommendation is that the state re-examine the proficient cut scores for the KPT to ensure that the scores are not inadvertently too low. Looking at data on proficiency rates, and comparing individual students' performance on the KPT and Spring AZELLA could be important activities to undertake in carefully re-examining the cut scores. Data analysis suggests raising the cut score by 3 to 4 points may decrease the likelihood that a large number of students will be identified as proficient on the KPT and non-proficient on the AZELLA, which is administered just a few months after the KPT.

Professional Development Related to the KPT. Additional online training focused on administering the KPT would also be beneficial for test administrators. The KPT training, in this case, could include more explanatory information on scoring each particular item. Furthermore, professional development related to Arizona College and Career Ready Standards, English language development, intercultural competency, and assessments in general may address issues related to the purpose of the KPT and the development of English language proficiency in Kindergarten.

SEI Professional Development. Although not directly related to the KPT administration, educators generally reported wanting additional instructional strategies to use in classrooms with ELLs. Overall, a need for more professional development focused on Kindergarten English language instruction was well-supported by the interviewees. The NCEO project team recommends that the Arizona Department of Education consider additional professional development for Kindergarten English language development in general, and SEI teaching in particular. In addition, the data analysis in Activity 6 indicates that students who were proficient on the KPT initially, but did not pass the AZELLA may need supplementary language instruction in the areas of reading and writing. Additional professional development to empower educators with strategies for teaching reading and writing in Kindergarten is recommended.

References

Arizona Department of Education (2012). AZELLA Arizona English Language Learner Assessment Reference Manual for 2012-2013. Retrieved from: http://www.azed.gov/standards-development-assessment/files/2012/08/reference-manual-revised-aug-15-2012.pdf

Arizona Education Statute, Title 15, Chapter 7, Article 3.1, "English Language Education for Children in Public Schools," §15-756 (2013).

Bailey, A. L. (Ed.). (2007). *The language demands of school: Putting academic language to the test*. New Haven, CT: Yale University Press.

Bailey, A. L., Butler, F. A., Laframenta, C., & Ong, C. (2004). *Towards the characterization of academic language in upper elementary science classrooms (CSE Report 621).* Los Angeles, CA: University of California Los Angeles, Center for Research on Evaluation, Standards, and Student Testing [CRESST], Center for the Study of Evaluation. Retrieved from http://www.cse.ucla.edu/products/reports/r621.pdf

Camara, W., & Lane, S. (2006). A historical perspective and current views on the Standards for Educational and Psychological Testing. *Educational Measurement: Issues and Practice*, *25*(3), 35-41.

Christopher, S., & Webb, N. (2013). *Alignment analysis of Arizona English Language Learners Assessment (AZELLA) and AZELLA Assessable Standards (2013), Kindergarten Placement Test and Stages I-V, Forms A and B.* Retrieved from http://www.azed.gov/standards-development-assessment/files/2013/10/alignment_analysis_azella_report_october_10_2013.pdf

Crocker, L. (1997). Editorial: The great validity debate. *Educational Measurement: Issues and Practice, 16*(2), 4.

Cronbach, L. J. (1989). Construct validation after thirty years. *Intelligence: Measurement, theory, and public policy*, 147-171.

Cronbach, L. J. (1988). Five perspectives on validity argument. *Test validity*, 3-17.

Fox, J., (2004). Test decisions over time: Tracking validity. *Language Testing, 21*, 437-465.

Green, D. (1998). Consequential aspects of the validity of achievement tests: A publisher's point of view. *Educational Measurement: Issues and Practice, 17*(2), 16-19.

Kane, M. (2002). Validating high-stakes testing programs. *Educational measurement: Issues and practice*, *21*(1), 31-41.

Lane, S., Parke, C., & Stone, C. (1998). A Framework for Evaluating the Consequences Programs of Assessment. *Educational Measurement: Issues and Practice, 17*(2), 24-28.

Parke, C. S., Lane, S., & Stone, C. A. (2006). Impact of a state performance assessment program in reading and writing. *Educational Research and Evaluation*, *12*(3), 239-269.

Lane, S., & Stone, C. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice, 21*(1), 23-30*.*

Linn, R. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice, 16*(2), 14-16

Linn, R. (1998). Partitioning responsibility for the evaluation of the consequences of assessment programs. *Educational Measurement: Issues and Practice, 17*(2), 28-30.

Marion, S. (2007). *A technical design and documentation workbook for assessments based on* modified achievement standards. Minneapolis, MN: National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/nceo/Teleconferences/ AAMASteleconferences/AAMASworkbook.pdf

Marion, S., & Pellegrino, J. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice, 25*(4), 47-57.

Marion, S., Quenemoen, R., & Kearns, J. (2006). *Introduction to seminars on inclusive assessments*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

McDonnell, L. M., & Choisser, C. (1997). *Testing and teaching: Local implementation of new state assessments (CSE Tech. Rep. No. 442)*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Moss, P. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice, 17*(2), 6-12.

Norris, J. (2008). *Validity evaluation in language assessment.* Frankfurt, Germany: Peter Lang.

Pellegrino, J., Chudowski, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Perie, M., & Forte, E. (2011) Developing a validity argument for assessments of students in the margins (p. 335- 376). In M. Russell, and M. Kavanaugh (Eds.), *Assessing students in the margin: Challenges, strategies, and techniques.* Charlotte, NC: Information Age Publishing, Inc.

Popham, J. (1997). Consequential validity: Right concern – wrong concept. *Educational Measurement: Issues and Practice, 16*(2), 9-13.

Reckase, M. (1998). Consequential validity from a test developer's perspective. *Educational Measurement: Issues and Practice, 17*(2), 13-16.

Scott, L. (2013). *Concurrent external validity study of Arizona's AZELLA Kindergarten placement test.* Retrieved from http://www.azed.gov/standards-development-assessment/files/2013/10/concurrent-external-validity-report_final.pdf

Shepard, L. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice, 16*(2), 5-24.

Yen, W. (1998). Investigating the consequential aspects of validity: Who is responsible and what should they do? *Educational Measurement: Issues and Practice, 17*(2), 5.

Appendices

State of Arizona
Department of Education
Office of English Language Acquisition Services

Primary Home Language Other Than English (PHLOTE)
Home Language Survey
(Effective April 4, 2011)

These questions are in compliance with Arizona Administrative Code, R7-2-306(B)(1), (2)(a-c).

Responses to these statements will be used to determine whether the student will be assessed for English Language Proficiency.

1. What is the primary language used in the home regardless of the language spoken by the student? _____

2. What is the language most often spoken by the student? _____

3. What is the language that the student first acquired? _____

Student Name _____ Student ID _____

Date of Birth _____ SAIS ID _____

Parent/Guardian Signature _____ Date _____

District or Charter _____

School _____

_____

Please provide a copy of the Home Language Survey to the ELL Coordinator/Main Contact on site.

In SAIS, please indicate the student's home or primary language.

1535 West Jefferson Street, Phoenix, Arizona 85007 • 602-542-0753 • www.azed.gov/oelas

Appendix B

Activity 1: Observation/Interview

Before test administration:

1. Type of building (public school, admin building, etc.): _____

2. Where was testing located in the building:

   _____

3. Child's parent present? □Yes □No

   If yes, did parent have any influence on the testing process? How?

   _____

4. Interpreter present? □Yes □No

5. Child's home language: □ Spanish   □ Navaho □ Arabic □ Vietnamese   □ Somali

   □ Other:_____

During test administration:

1. List the steps the test administrator followed in the testing session:

2. Did the test administrator follow testing protocol? □Yes □No

   If no, make notes about any protocol changes:

   _____
   _____

3. Did the test administrator repeat any of the test items? □Yes □No

   If yes, how many items were repeated? _____

4. What did you observe about the child's behavior during testing (check those that describe
   how the child behaved *most of the time*)?

a. Whole body movement: □ Sitting □ Standing □ Out of seat (moving around the room) □ Other _____

b. Face orientation: □ Facing test materials and administrator □ Turned away from administrator □ Other _____

c. Verbalization: □ Speaking/Responding in English □ Speaking/Responding in Non-English Language □ Non-Word Sounds □ None □ Other _____

d. Interaction with administrator: □ Responding on topic when prompted □ Responding off topic when prompted □ Interrupting □ None □ Other _____

5. Were there any unusual child-related or environmental factors that might have affected the results of testing (e.g.: fire alarm, phone call that interrupts testing, child needs to use the bathroom, etc.)? □Yes □No

If yes, describe: _____

After test administration:

1. Was the child's native language used during the assessment process? □Yes □No   If yes, by whom? □Child □Parent □Interpreter □Other: _____ If yes, what was the purpose? _____

2. Was the child able to complete all tasks in the time allotted? □Yes □No

3. Length of assessment (in minutes): _____

4. Were breaks offered? □Yes □No

5. The testing environment was: □Quiet □Noisy □Other:_____

6. Were distracting items present in the testing space (toys, other students, etc.)? □Yes □No If yes, describe: _____

7. Was equipment used (e.g., furniture) appropriate for the child, considering size and developmental level? □Yes □No If no, describe:_____

8. Was the child receiving special education services? □Yes □No

9. Once testing was completed, who entered the scores online? □Test Administrator □Support Staff □Other: _____ □Did not see

Comments:

_____
_____
_____

_____

_____

Post-Observation Follow-up Questions

Instructions for Observer/Interviewer: Test administrators might have different amounts of time to spend answering the following questions. Adjust prompts as needed to account for time allotment.

1. Describe your training experiences with the AZELLA Kindergarten placement test. When did you complete the training module? Did you receive any other training? Have you experienced answers given by children that you didn't know how to score/were not covered in training?

2. How many assessments did you give today? How many have you given total this school year? Do/did you complete assessments in multiple buildings?

3. Describe how scores are entered and how test administrators at your school ensure that data are entered correctly.

4. Have you experienced difficulties or challenges during the testing process?

   a. Have you ever experienced testing a child with unique needs? How did you respond?
   b. Which test items were easiest for the child? Which appeared to be the most difficult? Why?

5. How do you feel the testing session went for this child?

6. Do you have any additional comments?

Appendix C

Activity 2: Retrospective Interviews

Introduction

Hi I'm [name]. I'm visiting here from the National Center on Educational Outcomes (NCEO) at the University of Minnesota. NCEO is a technical assistance center that helps state departments of education improve their testing processes for students with disabilities, ELLs, and ELLs with disabilities.

Hand out business card for future questions and follow-ups

As you might know, NCEO was asked by the Arizona Department of Education to help figure out how the new AZELLA Kindergarten assessment is working. We're talking with test administrators to learn more about testing procedures, any unanticipated issues that are coming up during testing, and additional needs of test administrators. I/We appreciate that you are taking the time to talk to me/us today. I/We are going to take notes on your responses but I am/we are not recording your name. The state department of education staff helped us find you as a potential person to interview so they will know we talked to you, but they will not be able to connect responses to a specific person or school. The focus of our interviews is on improving the <u>test</u>, not evaluating your work as a test administrator.

I/We want to make sure you are aware that I/we have completed certification to administer the AZELLA Kindergarten Placement test and I/we understand the confidential nature of the test items. The state has given me/us permission to be able to discuss the individual items today.

Do you have any questions before we begin?

Our interview today has two parts. It should last about 30-45 minutes. First I/we will ask you some questions about your overall experiences with test administration and the testing process. Second I/we will ask you about your opinions of the testing.

A. Test Administrator Background Information

1. Tell me how the test administration went this fall (AZELLA Kindergarten placement test, NOT other AZELLA tests).

   - *PROMPT*: When did you give the test?
   - *PROMPT:* How many kids did you test?
   - *PROMPT:* Did you enter the scores yourself? When? When did you get the results back? Were they helpful?
   - *PROMPT:* When did you complete the online training? What did you think of the training?

2. About how many students did you test this fall? How many tests per week are you giving now? Generally, how did the students do? Were there any placement changes based on the test results?

3. What is your understanding about the purpose of this test? There is no right or wrong answer to this question.

B. Information about the testing process

1. Describe a typical testing experience with a child. What do you usually do first? What do you do next?

    PROMPTS: Did they know the child? Were they able to establish rapport? What was the setting like (location, etc.)?

    [ ]

2. Going back to the tests you gave this fall, think of a student/students for whom the test was easy.

    -   How would you describe the testing process?
    -   Did the student produce enough language for you to have confidence in the language proficiency determination?
    -   Do you think that the performance of the student was an accurate representation of the student's actual language ability? Why or why not?

    [ ]

3. Now think of a student/students who had difficulties with the assessment.

- How would you describe the testing process?
- Did the student produce enough language for you to have confidence in the language proficiency determination?
- Do you think that the performance of the student was an accurate representation of the student's actual language ability? Why or why not?

4. Did you test any students with disabilities? If so, what was that like?

5. Were there any children for whom the proficiency determination surprised you?
- Tell me a little about that testing situation
- What made the result surprising for you?

C. We've been talking about these example students. Now I'd like to ask you more broadly about the test and your opinions about it. I'd like to remind you that we are not breaking the security of the test by talking about the individual items.

    1. Overall, which test items seem to be the easiest for the children? Which appeared to be the most difficult? Why?

    2. Were the questions relevant?

    3. Do you think there are enough questions on the test for you to have confidence in the language proficiency determination? Tell me more about your answer.

    4. Do you think this test appropriately identifies students in need of SEI programming? Tell me more about that.

5. Do you have any other comments regarding your experiences administering the AZELLA Kindergarten Test?

Thank you again for talking with us. The information you have provided will be very useful to helping the Arizona Department of Education make decisions about the testing process. We will be summarizing the results of all our interviews and sharing them with ADE in the near future. Your responses will be included but you will not be identifiable in the data.

Appendix D

Activity 4: Case Study Data Collection Tools

SCHOOL STRUCTURE & DEMOGRAPICS INTERVIEW – (ESL COORDINATOR AND PRINCIPAL)

Arizona Case Study

Hi, I am NAME. And this is NAME. We are visiting here from the National Center on Educational Outcomes (NCEO) at the University of Minnesota. NCEO is a technical assistance center that helps state departments of education improve their testing processes for, ELLs, students with disabilities and ELLs with disabilities.

Here is Dr. Laurene Christensen's card in case you need to contact us in the future about this project.

As you might know, NCEO was asked by the Arizona Department of Education to help figure out how the AZELLA Kindergarten assessment is working. Before we worked for NCEO, we worked as (classroom experience/ELL mentioned here).

We are interested in learning how ELLs are doing after taking the test and what language is used in the classroom. Our project team has been able to observe some AZELLA kinder assessments; we've interviewed people about their impressions of the assessment; we've observed the AZELLA 3 administrations. Right now, we are doing some case studies to better understand the effect of this Kindergarten assessment on a district.

We appreciate that you are taking the time to meet with us today. We are going to take notes on your responses but we are not recording your name. In any reports that we release, we will talk about your school and district only in general terms. The Arizona Department of Education staff helped us find your school as a site for this project but they will not be able to connect responses to a specific person or school. .

Do you have any questions before we begin?

1. **Tell me a little about your school**
   a. Does your school have a particular mission or focus?

   b. How many students are enrolled? In what grades?

   c. How many students are enrolled in your district? How many schools are in the district? What grade levels do they serve?

   d. How would you describe the location of your school/district? (urban, suburban, rural)

   e. Who are your students? (e.g., racial/ethnic makeup, home languages, SES)

   f. What type of ESL programming do you offer?

   g. How many teachers do you have? How many teach in SEI classrooms?

   h. How many ELLs have you typically had in the past?

   i. How many ELLs did you serve this year?

2. **How is your school day structured? (e.g., 90 minute blocks on a rotating schedule)**

3. **Has your school made the requirements for Adequate Yearly Progress (AYP) this year? Tell me more about that.**

4. **Is there other information it would be important for us to understand about your school as we examine the effects of implementing the AZELLA pre-Kindergarten assessment this year?**

At the conclusion:

Thank you again for taking the time to meet with us. If you have any additional questions or comments, please feel free to connect with us while we are here this week or follow up with Dr. Christensen via email.

As a token of our appreciation, here's a little gift from Minnesota.

Observation of Academic Language in Arizona Kindergarten Classrooms

Classroom Information (complete with teacher during interview)

Date: _____   Observer:_____

Number of students:_____(____boys and ____ girls)

Number/Proportion of ELL students in class:_____

English proficiency level of the ELLs (i.e., ELD level) :_____

Number of students with disabilities: _____

Other adults present in room: _____

Type of classroom: Mainstream K/ SEI Kindergarten/other

Duration of observation in minutes:_____

Topic of lesson(s):_____

Unexpected events: _____

Number of years teacher has been teaching overall: _____

Number of years teacher has taught in this school: _____

Does the teacher have SEI endorsement: _____

Adapted from the Academic Language Exposure Checklist (ALEC) by Bailey, Butler, LaFramenta, & Ong (2004)

Part 1: Classroom Activities

1.1 Describe main activities observed (e.g., group instruction, individual problem solving, silent reading, small group collaborative work)

1.1a List any materials used (i.e., books, manipulatives, web site, work packet)

1.2 Different groupings of interlocutors addressed by teacher and/or teacher aides:
☐Whole Class   ☐Small Groups   ☐Pairs   ☐Individuals

Examples:
1.3 Different groupings of interlocutors addressed by students
☐Teacher   ☐Whole Class   ☐Small Groups   ☐Pairs   ☐Individuals

Examples:

2.1 Language demonstrated by teacher

Oral Language (check all that apply)

Language Function	Number of Times	Examples
☐Explanation		
☐Description		
☐Comparison		
☐Clarification		
☐Direction		
☐Other (Describe)		

Listening (Check all that apply)

Type of Response	Number of Times	Examples
☐ Verbal (e.g., asks another question, gives oral feedback, uses rewarding words)		
☐ Nonverbal (e.g., nods in agreement, uses facial expression)		

Reading (check all that apply)

☐ Oral

☐ Silent

What was the teacher reading?

Writing (check all that apply)

Type of Text Written	Number of Times	Examples
☐ Directions for activities		
☐ Questions/Problems		
☐ Answers to questions/problems		
☐ Stories (e.g., group-constructed story)		
☐ Other		

2.2 Does teacher ask similar questions (e.g., paraphrase) in different ways?

☐Yes   ☐No

Examples:

2.3 How does the teacher introduce new vocabulary? (Check all that apply)

	Number of Times	Examples
□ Gives definition		
□ Uses synonyms		
□ Gives examples		
□ Repeats words in context		
□ Asks questions to elicit definition		
□ Uses picture or visual		
□ Other		

Sample of nonspecialized words teacher used:

_____
_____
_____
_____
_____

Specialized words teacher used:

_____
_____
_____
_____
_____

3. Students' Language

3.1 Language produced by students

Oral Language (check all that apply)

	Number of Times	Examples
☐ Explanation of thinking or answer		
☐ Short answer to question (*E.g., T: What color is this? S: Blue)*		
☐ Description		
☐ Comparison		
☐ Justification of answer		
☐ Other (describe)		

Listening (check all that apply)

Type of Response	Number of Times	Examples
□ Verbal (e.g., asks another question, gives oral feedback, uses rewarding words)		
□ Nonverbal (e.g., nods in agreement, uses facial expression)		

Reading (check all that apply)

□ Oral

□ Silent

What were students reading?

Writing (check all that apply)

Type of Text Written	Number of Times	Examples
□ Directions for activities		
□ Questions/Problems		
□ Answers to questions/problems		
□ Stories		

□ Other		

3.2 Did students echo the teacher's syntactic structures and vocabulary in oral or written speech?

□ Yes □ No

Examples:

3.3 Students' use of academic vocabulary

Sample of nonspecialized words student used:

_____
_____
_____
_____
_____

Specialized words student used:

_____
_____
_____
_____
_____

3.4 How do ELLs/FEP students signal their lack of understanding?

Signal	Number of Times Observed	Same Student Each Time?	Examples of signal
☐ Ask for repetition of directions or information		☐ Yes ☐ No	
☐ Ask for another example		☐ Yes ☐ No	
☐ Ask for additional definition		☐ Yes ☐ No	
☐ Ask peers for help		☐ Yes ☐ No	
☐ Overtly state lack of understanding		☐ Yes ☐ No	
☐ Other _____		☐ Yes ☐ No	
☐ No observable signals			

4. Fluent English Speaker vs. ELL practices (Mainstream classrooms only)

4.1 Does the teacher use different language for addressing ELLs vs. fluent English speakers?

	For whom?	**Examples**
□ Uses students native language	□ for ELLs □ for fluent English speakers	
□ Simplifies explanations/provides further explanation	□ for ELLs □ for fluent English speakers	
□ Uses additional examples/counter-examples	□ for ELLs □ for fluent English speakers	
□ Changes discourse (explicit vs. implicit)	□ for ELLs □ for fluent English speakers	
□ Other (describe)	□ for ELLs □ for fluent English speakers	

Kindergarten Teacher Interview

Arizona case study

Hi, I am NAME. And this is NAME. We are visiting here from the National Center on Educational Outcomes (NCEO) at the University of Minnesota. NCEO is a technical assistance center that helps state departments of education improve their testing processes for, ELLs, students with disabilities and ELLs with disabilities.

Here is Dr. Laurene Christensen's card in case you need to contact us in the future about this project.

As you might know, NCEO was asked by the Arizona Department of Education to help figure out how the AZELLA Kindergarten assessment is working. Before we worked for NCEO, we worked as (classroom experience/ELL mentioned here).

We are interested in learning how ELLs are doing after taking the test and what language is used in the classroom. Our project team has been able to observe some AZELLA kinder assessments; we've interviewed people about their impressions of the assessment; we've observed the AZELLA 3 administrations. Right now, we are doing some case studies to better understand the effect of this Kindergarten assessment on a district.

We appreciate that you are taking the time to meet with us today. We are going to take notes on your responses but we are not recording your name. In any reports that we release, we will talk about your school and district only in general terms. The Arizona Department of Education staff helped us find your school as a site for this project but they will not be able to connect responses to a specific person or school.

Do you have any questions before we begin?

➢ *Do the first page of classroom observation form together (e.g., demographics).*

1. **What were your instructional goals for the lesson we observed?**
   PROMPTS: How representative of your typical instruction during the year was the lesson we observed? Is there anything about your classroom that we should know to understand the lesson? Did the students behave about like they usually do?

2. **What kinds of units have you taught this year? To what degree has your teaching aligned with content/ELP standards for Kindergarten?**

   PROMPTS: Which standards did you cover in the most detail? Which standards did you cover in the least detail?

3. **Think back to the kids who took the AZELLA Kindergarten placement test in the fall (either ELLs or Fluent English Proficient [FEP] students). How are these students doing in your class overall?**

   PROMPTS: Are they making progress in learning the instructional content? What kinds of data do you use to help you figure out if they are making progress?

4. **What English skills do you think students need to have to be successful in your Kindergarten classroom?**

   PROMPTS: How many of your ELLs (Fluent English Proficient [FEP] students) had these skills at the beginning of the school year? [For mainstream teachers] How many of your other students had these skills at the beginning of the school year?

5. **What English skills do you specifically teach during the school year?**

PROMPTS: How are your ELLs (FEP students) doing in learning these skills? [For mainstream teachers] How are other students doing in learning these skills?

6. **How do you scaffold instruction for ELLs? What do you do to help them understand?**

   PROMPT: Which of these things have been most successful? Which have been least successful? Tell me more about that.

7. **How well do you think the SEI certification prepared you to be an effective teacher of ELLs (FEP students)?**

8. **What types of opportunities exist for you to collaborate with ESL/mainstream teachers when you are planning instruction for ELLs?**

9. **How has the implementation of the AZELLA Kindergarten placement test affected or not affected your job as a Kindergarten teacher this year?**

   PROMPT: Have you used the results to make any instructional decisions? How have you communicated with parents about the Kindergarten placement test results? How have they responded?

10. **Do you have any other comments or questions that would help us understand the ways that you see academic English being used in your classroom and how that relates to the AZELLA placement test?**

At the conclusion:

Thank you again for taking the time to meet with us. If you have any additional questions or comments, please feel free to connect with us while we are here this week or follow up with Dr. Christensen via email.

As a token of our appreciation, here's a little gift from Minnesota.

Principal Interview

Arizona Case Studies

Hi, I am NAME. And this is NAME. We are visiting here from the National Center on Educational Outcomes (NCEO) at the University of Minnesota. NCEO is a technical assistance center that helps state departments of education improve their testing processes for, ELLs, students with disabilities and ELLs with disabilities.

Here is Dr. Laurene Christensen's card in case you need to contact us in the future about this project.

As you might know, NCEO was asked by the Arizona Department of Education to help figure out how the AZELLA Kindergarten assessment is working. Before we worked for NCEO, we worked as (classroom experience/ELL mentioned here).

We are interested in learning how ELLs are doing after taking the test and what language is used in the classroom. Our project team has been able to observe some AZELLA kinder assessments; we've interviewed people about their impressions of the assessment; we've observed the AZELLA 3 adminstrations. Right now, we are doing some case studies to better understand the effect of this Kindergarten assessment on a district.

We appreciate that you are taking the time to meet with us today. We are going to take notes on your responses but we are not recording your name. In any reports that we release, we will talk about your school and district only in general terms. The Arizona Department of Education staff helped us find your school as a site for this project but they will not be able to connect responses to a specific person or school.

Do you have any questions before we begin?

1. **How did you find out about the AZELLA Kindergarten placement test?**
   PROMPTS: Who communicated this information to you? When did you learn about the test?

2. **How did the AZELLA Kindergarten placement test go this fall?**
   PROMPTS: How many students did you test? What went well? What challenges did you experience?

3. **What types of decisions did your school/district make using the results of the AZELLA Kindergarten placement test?**

    PROMPTS: Who was involved in making these decisions? Did you reclassify students who were already placed at the start of the school year? Do you think any of the students who initially passed the placement test will get reclassified for next year based on the AZELLA 3 results?

4. **How has the school/district communicated with parents about the assessment process and the decisions made based on the AZELLA Kindergarten placement test results?**

    PROMPT: In what form was the communication (oral, written)? Who was responsible for the communication? How have parents responded?

5. **How has the implementation of the Kindergarten placement test affected your job?**
   PROMPTS: What effect, if any, has it had on staffing and hiring? What effect, if any, has it had on planning and providing staff development and training? What effect, if any, has it had on schedule planning?

6. **What opportunities, if any, do ESL and mainstream Kindergarten teachers have to collaborate in planning instruction for ELLS?**
   PROMPTS: If collaboration isn't happening now, what kinds of resources or structures would it take to support more collaboration? (If collaboration is happening now) Are there ways to make this collaboration more effective?

7. **What could be done to ensure that AZELLA Kindergarten placement test results are as useful as possible for educational decision-making?**

8. **Other comments or questions?**

At the conclusion:

Thank you again for taking the time to meet with us. If you have any additional questions or comments, please feel free to connect with us while we are here this week or follow up with Dr. Christensen via email. As a token of our appreciation, here's a little gift from Minnesota.

ESL Coordinator Interview

Arizona Case Studies

Hi, I am NAME. And this is NAME. We are visiting here from the National Center on Educational Outcomes (NCEO) at the University of Minnesota. NCEO is a technical assistance center that helps state departments of education improve their testing processes for, ELLs, students with disabilities and ELLs with disabilities.

Here is Dr. Laurene Christensen's card in case you need to contact us in the future about this project.

As you might know, NCEO was asked by the Arizona Department of Education to help figure out how the AZELLA Kindergarten assessment is working. Before we worked for NCEO, we worked as (classroom experience/ELL mentioned here).

We are interested in learning how ELLs are doing after taking the test and what language is used in the classroom. Our project team has been able to observe some AZELLA kinder assessments; we've interviewed people about their impressions of the assessment; we've observed the AZELLA 3 administrations. Right now, we are doing some case studies to better understand the effect of this Kindergarten assessment on a district.

We appreciate that you are taking the time to meet with us today. We are going to take notes on your responses but we are not recording your name. In any reports that we release, we will talk about your school and district only in general terms. The Arizona Department of Education staff helped us find your school as a site for this project but they will not be able to connect responses to a specific person or school. .

Do you have any questions before we begin?

1. **How did the AZELLA Kindergarten placement test go this fall?**
   PROMPTS: How many students did you test? What went well? What challenges did you experience?

2. **What types of decisions did your school/district make using the results of the AZELLA Kindergarten placement test?**
   PROMPTS: Who was involved in making these decisions? Did you reclassify students who were already placed at the start of the school year? Do you think any of the students who initially passed the placement test will get reclassified for next year based on the AZELLA 3 results?

3. How do you think the language skills assessed by the Kindergarten placement test compare to the language skills students need to have to participate in the Kindergarten content standards?

    PROMPTS: Are there additional language skills students need to successfully participate in Kindergarten classroom routines and activities? Tell me about these.

4. How has the school/district communicated with parents about the assessment process and the decisions made based on the AZELLA Kindergarten placement test results?

    PROMPT: In what form was the communication (oral, written)? How have parents responded?

5. **How has the implementation of the Kindergarten placement test affected your job?**
   PROMPTS: What are your primary duties as ESL coordinator? Are there any responsibilities you have now that you did not have before the testing began? What effect, if any, has the testing had on staff training and development efforts that you plan or provide? What effect, if any, has the test had on curriculum planning?

6. **What opportunities, if any, do ESL and mainstream Kindergarten teachers have to collaborate in planning instruction for ELLS?**
   PROMPTS: If collaboration isn't happening now, what kinds of resources or structures would it take to support more collaboration? (If collaboration is happening now) Are there ways to make this collaboration more effective?

7. **What could be done to ensure that AZELLA Kindergarten placement test results are as useful as possible for educational decision-making?**

8. Other comments or questions?

At the conclusion:

Thank you again for taking the time to meet with us. If you have any additional questions or comments, please feel free to connect with us while we are here this week or follow up with Dr. Christensen via email.

As a token of our appreciation, here's a little gift from Minnesota.

School/District Assessment and Evaluation Coordinator Interview

AZ Case Study

Hi, I am NAME. And this is NAME. We are visiting here from the National Center on Educational Outcomes (NCEO) at the University of Minnesota. NCEO is a technical assistance center that helps state departments of education improve their testing processes for, ELLs, students with disabilities and ELLs with disabilities.

Here is Dr. Laurene Christensen's card in case you need to contact us in the future about this project.

As you might know, NCEO was asked by the Arizona Department of Education to help figure out how the AZELLA Kindergarten assessment is working. Before we worked for NCEO, we worked as (classroom experience/ELL mentioned here).

We are interested in learning how ELLs are doing after taking the test and what language is used in the classroom. Our project team has been able to observe some AZELLA kinder assessments; we've interviewed people about their impressions of the assessment; we've observed the AZELLA 3 administrations. Right now, we are doing some case studies to better understand the effect of this Kindergarten assessment on a district.

We appreciate that you are taking the time to meet with us today. We are going to take notes on your responses but we are not recording your name. In any reports that we release, we will talk about your school and district only in general terms. The Arizona Department of Education staff helped us find your school as a site for this project but they will not be able to connect responses to a specific person or school. .

Do you have any questions before we begin?

1. **What types of standardized reading assessment data do you/does your school or district collect?**

   PROMPTS: What tests are used? When are they given? Who takes these tests? Are they norm or criterion-referenced? Which ones are given to Kindergarteners?

2. **How are data on these assessments disaggregated?**

   PROMPTS: Are the scores broken out for ELLS? Are there any disaggregated data on Kindergarten ELLs and initially fluent English proficient (IFEP) students that you could share with us?

3. **How do students perform on these other reading assessments?**
   PROMPTS: How do initially fluent English proficient (IFEP) students perform on these other reading measures? How do current ELLs perform on these other measures? Do any of the data show the performance of ELLS and IFEP students in Kindergarten?

4. **Do you track the change in test scores across years for a cohort of students?**
   PROMPTS: What does the pattern of change typically look like for IFEP students? What does the pattern of change typically look like for ELLs?

5. **Do you collect any other data that would be useful in determining how the AZELLA Kindergarten placement test is functioning?**

6. **Other comments/questions**

At the conclusion:

Thank you again for taking the time to meet with us. If you have any additional questions or comments, please feel free to connect with us while we are here this week or follow up with Dr. Christensen via email.

As a token of our appreciation, here's a little gift from Minnesota.

Appendix E

Activity 5: Principal Survey

AZELLA School Principal Survey

Does your school serve English language learners?

○ Yes

○ No

| ⁋ | 0 |

1. In what type of district do you work?

○ Urban

○ Suburban

○ Small town

○ Rural

○ Other: [   ]

2. In what type of school do you work?

○ District

○ Charter

○ Other: [   ]

3. What grades does your school serve?

Check all that apply

☐ K

☐ 1

☐ 2

☐ 3

☐ 4

☐ 5

☐ 6

☐ 7

☐ 8

☐ 9

☐ 10

☐ 11

☐ 12

4. What percentage of your total school population was identified as English language learners last year?

[   ]

5. What percentage of your total school population comes from a home where a language other than English is spoken?

[   ]

6. What type of programming does your school offer for ELLs?

○ Structured English Immersion (SEI)

○ Individual Language Learner Plan (ILLP)

○ Bilingual

○ Other: [   ]

7. How many teachers were assigned solely to ELD (SEI or Bilingual) classes in your school during 2012-13?

[   ]

8. How many staffers other than teachers were assigned solely to ELD (SEI or Bilingual) classes in your school during 2012-13?

[   ]

9. How many students who were identified as "Initially Fluent English Proficient" (IFEP) on the Kindergarten Placement Test in the fall of 2012 and were not proficient on the Spring 2013 AZELLA Reassessment, were placed in ELD (SEI or Bilingual) classrooms in 2013-14?

[_____]

10. How many students who were identified as "Initially Fluent English Proficient" (IFEP) on the Kindergarten Placement Test in the fall of 2012 and were not proficient on the Spring 2013 AZELLA Reassessment, will remain in Mainstream classrooms in 2013-14?

[_____]

11. For students identified as "Initially Fluent English Proficient" (IFEP) who did not score proficient on the Spring 2013 AZELLA Reassessment, and will remain in Mainstream classrooms, which of the following impacted your decision?

Check all that apply.

☐ Students have the skills needed for instruction in Mainstream classrooms

☐ Scores arrived after classrooms were assigned and moving students would be disruptive

☐ Meeting with parents was not practical

☐ Other: [_____]

12. Between what dates do you expect to have given the majority of Kindergarten Placement Tests during the 2013-14 school year?

[_____]

13. Test Administration of the Spring 2013 AZELLA Reassessment

	Strongly agree	Mildly agree	Neither agree nor disagree	Mildly disagree	Strongly disagree	I do not know
My school had the resources needed to administer the Spring 2013 AZELLA Reassessment.	○	○	○	○	○	○
The amount of time required to give the Spring 2013 AZELLA Reassessment was reasonable for me.	○	○	○	○	○	○
The amount of time needed to take the Spring 2013 AZELLA	○	○	○	○	○	○

	Strongly agree	Mildly agree	Neither agree nor disagree	Mildly disagree	Strongly disagree	I do not know
Reassessment was reasonable for the students.						
While taking the Spring 2013 AZELLA Reassessment, the students had the ability to use accommodations they required for a documented disability.	○	○	○	○	○	○
Staff knew how to provide any needed accommodations for the Spring 2013 AZELLA Reassessment.	○	○	○	○	○	○
Standardized test administration procedures were followed by all test administrators.	○	○	○	○	○	○

14. Please provide any additional comments about the Spring 2013 AZELLA Reassessment administration.

15. Score Use

	Strongly agree	Mildly agree	Neither agree nor disagree	Mildly disagree	Strongly disagree	I do not know
The students' Spring 2013 AZELLA Reassessment	○	○	○	○	○	○

102

	Strongly agree	Mildly agree	Neither agree nor disagree	Mildly disagree	Strongly disagree	I do not know
results are an accurate representation of their English proficiency skills.						
The students' Spring 2013 AZELLA Reassessment results are an accurate representation of their ability to succeed in the general education curriculum without special (or additional) language services.	○	○	○	○	○	○
The same students who were identified as ELLs in Kindergarten during Fall 2012 were identified by the Spring 2013 AZELLA Reassessment.	○	○	○	○	○	○
The scores and other performance information resulting from the Spring 2013 AZELLA Reassessment are meaningful to educators.	○	○	○	○	○	○
The scores and other performance information resulting from the Kindergarten Placement Test are meaningful to	○	○	○	○	○	○

	Strongly agree	Mildly agree	Neither agree nor disagree	Mildly disagree	Strongly disagree	I do not know
educators.						

16. Please provide any additional comments about the Kindergarten Placement Test and Spring 2013 AZELLA Reassessment score use.

17. Implications of the Test

|	Strongly agree	Mildly agree	Neither agree nor disagree	Mildly disagree	Strongly disagree	I do not know
The Spring 2013 AZELLA Reassessment results have had an effect on teacher hiring for the 2013-14 school year.	○	○	○	○	○	○
There are enough training opportunities available to prepare staff who will teach in ELD (SEI or Bilingual) classrooms for the first time in 2013-14.	○	○	○	○	○	○
The Spring 2013 AZELLA Reassessment results positively affected my school's A-F School Accountability Letter Grade.	○	○	○	○	○	○

18. Please provide any additional comments about implications of the Spring 2013 AZELLA Reassessment.

104
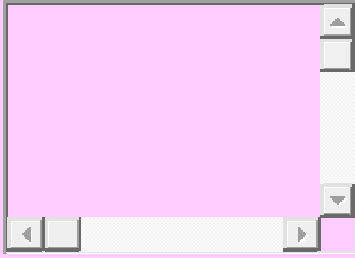
19. I have sufficient numbers of ELD (SEI or Bilingual) teachers for the 2013-14 school year.

○ Yes

○ No

○ Other: [_____]

20. I had to ask some general education teachers to teach in ELD (SEI or Bilingual) classrooms in the 2013-14 school year.

○ Yes

○ No

○ Other: [_____]

21. I have received Title III funding for all of the students in ELD (SEI or Bilingual) classrooms in 2013-14.

○ Yes

○ No

○ Other: [_____]

22. The Spring 2013 AZELLA Reassessment results are being communicated to parents.

○ Yes

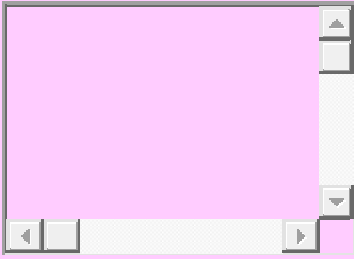○ No

○ Other: [_____]

23. Parents have asked questions or raised concerns about the results of the Spring 2013 AZELLA Reassessment.

○ Yes

○ No

○ Other: [_____]

24. How did you communicate the Spring 2013 AZELLA Reassessment results to those parents whose children did not pass the test after originally passing the Kindergarten Placement Test?

25. Do you have any other comments?

[[[,1007447520,["Ye   0,1