Winsight™
Created by ETS. Delivered by Questar.

# *Winsight*® Assessment System Preliminary Theory of Action

## *Abbreviated Version*

E. Caroline Wylie, Research Director, Student & Teacher Research, ETS

The *Winsight*® Assessment System integrates summative, interim and formative assessment components, initially focused on mathematics and English Language Arts (ELA) in grades 3–8 and high school. The theory of action sets out how the Winsight Assessment System is intended to improve teaching and learning when used in combination with professional supports (e.g., resources to support teachers' understanding of learning progressions and key practices, assessment literacy and other aspects of practice related to use of the Winsight Assessment System) in a sustained manner. As the program develops and matures, Winsight-specific evidence will be sought to support and/or refine the theory of action.

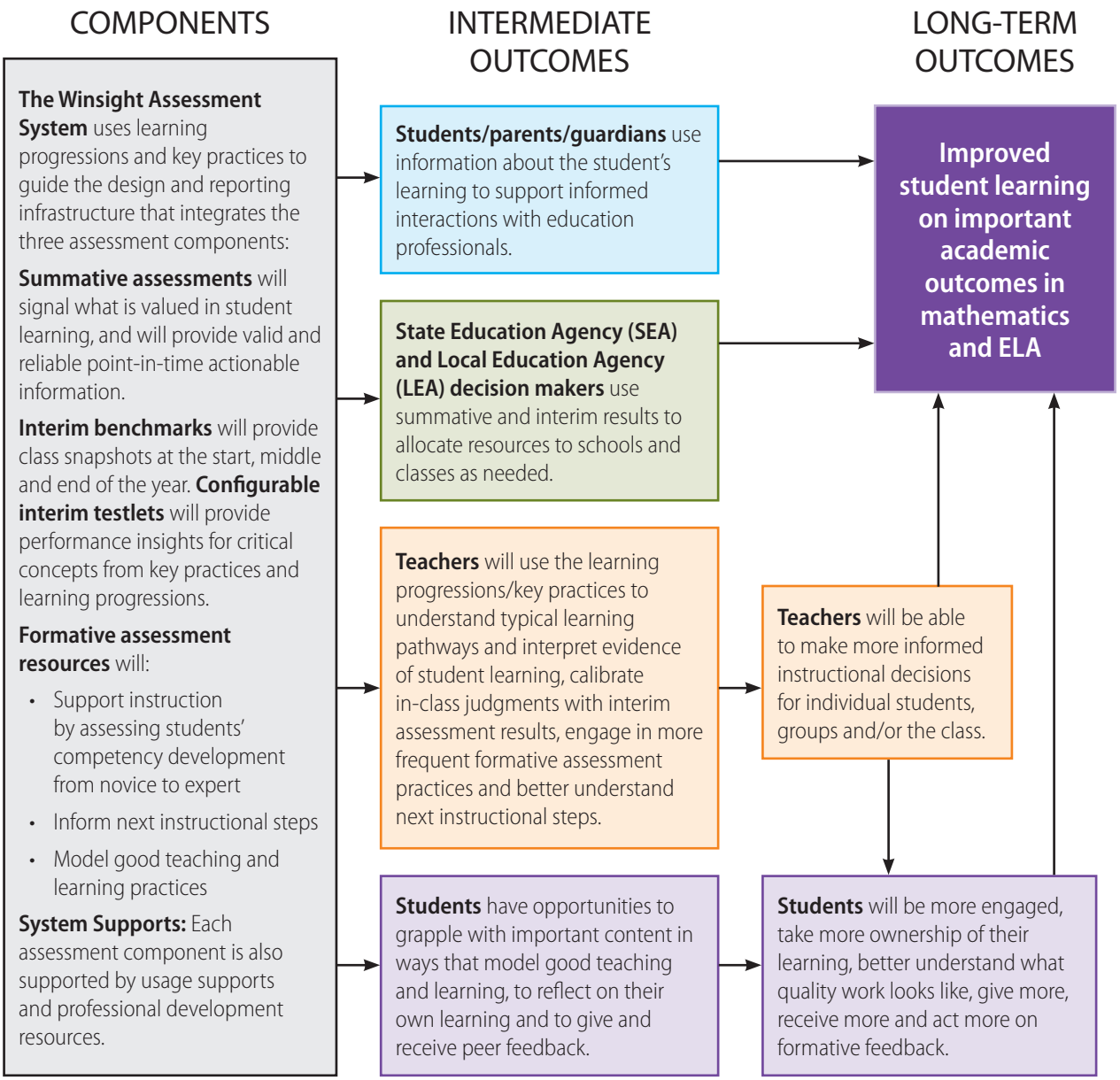
**Components of the Winsight Assessment System**

The Winsight Assessment System will combine four assessment components into one online system aligned to state standards, as well as a unique combination of learning progressions across the system. Learning progressions (LPs) and key practices (KPs) are theories of students' competency development that articulate how learning develops from a novice understanding or practice of a competency to a more sophisticated understanding or practice.

1. The **summative assessment** is primarily intended to measure students' proficiency with respect to each state's standards in ELA and mathematics in the initial launch (with science anticipated for future versions) for accountability purposes. At launch, the summative assessment will be a conventional linear test, after which it will use a multistage adaptive approach to tailor the assessment to student ability. The results (total scaled score, sub-scores, proficiency levels, for all students and for sub-groups) will be available in an interactive reporting system that can be used by state, district and school administrators to examine student progress toward college and career readiness in mathematics and ELA.
2. **Interim assessments** are intended to be used by teachers for instructional modification and by administrators for monitoring student learning between annual summative administrations. Interim *benchmark* assessments provide snapshots at the start, middle and end of the year, and are intended to be predictive of student performance on the summative assessment. The interim *testlet* assessments will be developed to target critical aspects of standards, replacing what is often referred to in schools as *common assessments*. They can be used at flexible intervals as best determined by the needs of the classroom teacher, such as at the start or end of a unit focused on proportional reasoning in mathematics or argumentation in ELA. They will provide results that highlight progress on critical LPs and KPs.
3. The **formative assessment** component of the system is a set of resources that model good teaching and learning practices to support formative assessment processes and ongoing instruction, and to track student progress within tasks, assessing students' competency development along a continuum of development from novice to expert. In the final product, there may be a range of formative resources from quick checks-for-understanding to more extensive scenario-based tasks that provide more scaffolded learning opportunities. Reports in the formative setting will focus on providing finer-grained information (compared to the testlets) to help teachers make sense of student responses in light of the standards, LPs and KPs, and will provide guidance to the next instructional steps. While testlets support periodic assessment, the formative assessment resources will support teachers' and students' ongoing daily assessment of learning.

4. While not an assessment component, an important aspect of the Winsight Assessment System will be the **system supports**. System supports are intended to encompass a wide range of resources beyond the manuals and online resources that support test administrations. System supports also include resources that support assessment literacy and that help teachers understand and make use of the LPs and KPs, both as part of Winsight assessment reports and as teaching and learning supports. We plan to provide resources that will model good teaching and learning practices, supported by just-in-time professional development opportunities through webinars, online videos and resources for school-based professional learning communities.

The role of the logic model for the Winsight Assessment System is to represent the set of claims that we are building the system to achieve. We can look to existing research studies to support the logic and plausibility of these claims, and to identify gaps where we may wish to target initial research. Below we present a high-level logic model and describe sets of claims, organized by the key users of the system. A more detailed version is available at www.ets.org (Wylie, 2017).



## Claim 1: Impact on students and parents/guardians

*When students and parents/guardians have access to information about student learning from the Winsight Assessment System components, they are able to have informed interactions with education professionals. Improved communication between parents/guardians and education professionals has a positive impact on student learning outcomes.*

The claims associated with the blue box in the theory of action graphic argue that when parents and guardians have access to appropriate information (both student results and guides to interpreting the various forms of information) about student learning, they are able to have more informed interactions with school-based education professionals and that having parents/guardians more engaged leads to improved student learning outcomes. Research shows that parents of color and parents with lower income levels were more likely to value the information coming from summative accountability assessments (Tompson, Benz, & Agiesta, 2013). Communicating information about student results in ways that all parents/guardians can understand and take action on is important (Zapata-Rivera, Nabors-Olah, Vezzu, Biggers, Leusner, & Bertling, 2013). Research on score reports emphasizes the importance of providing information in a valid, reliable and meaningful way, with support for parents to engage with school personnel (Munk & Bursuck, 2001). There is also research evidence that demonstrates that when parents are generally more involved with their children's education, there is a positive impact on student learning (Jeynes, 2005; Jeynes, 2007; Shute, Hansen, Underwood, & Razzouk, 2011). Less is known about how to present assessment information to students in a way that encourages continued or improved engagement with learning. It will be important to study whether and how the Winsight assessment reports are able to support parent-teacher involvement in meaningful ways, along with investigations on the nature of other supports and structures for both parents and students.

## Claim 2: Impact on school, district and state-level administrators

*When State Education Agency (SEA) and Local Education Agency (LEA) decision makers have access to valid and reliable information about student learning from the Winsight summative and interim assessments, they are able to allocate resources more appropriately at the school and classroom level, which enhances student learning.*

The claims associated with the green box in the theory of action graphic argue that when SEA and LEA decision makers have access to reliable and valid summative and interim assessment information, they are better able to allocate supports and resources to schools and classes as needed. As part of the development of these components, there will be a constant focus on identifying information for school, district and state decision makers that will support actionable decision making through the use of critical LPs/KPs to provide information that goes beyond proficient/not proficient. Many SEAs have implemented processes to group schools according to achievement on accountability assessments and to provide resources to those schools and districts. For example, North Carolina identifies two types of schools — Priority and Focus schools — and provides professional development for school leaders, resources to support family and community engagement, instructional coaches, and technology supports (http://dst.ncdpi.wikispaces.net) for each group. As a second example, California has implemented a process of Program Improvement for schools and districts and has identified a system of supports that can be accessed for schools designated as low-performing (http://www.cde.ca.gov/sp/sw/ss/index.asp). Similar decision-making processes and supports can be found for most states, although impact on student learning is more difficult to find. More research has been done on the conditions that are necessary for administrators to use student data than on directly looking at those impacts (e.g., Honig & Coburn, 2007; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006). However, in the context of specific sub-groups, researchers such as Ysseldyke, Dennision, & Nelson (2004) noted that the effect of the IDEA mandate has been to "raise the bar" for students with disabilities by using summative assessments to identify curriculum areas that need a targeted focus in a particular grade level, and to guide curriculum emphases or revisions.

## Claim 3: Impact on teachers

> *When teachers have access to assessment information from multiple Winsight components about students, they use the LPs/KPs to interpret evidence of student learning, calibrate in-class judgments with interim assessment results, engage in more frequent formative assessment practices and better understand next instructional steps. When teachers better understand student competency development and use resources from the Winsight formative component to support their formative assessment practices, they make more informed next instructional decisions to tailor instruction to the specific needs of individual students, groups of students or the whole class based on evidence, which improves student learning.*

The claims associated with the orange boxes in the theory of action graphic lay out the intermediate impacts on teachers as a result of the use of the Winsight Assessment System. These impacts include providing access to relevant student information and supporting the use of formative assessment practices, which lead to better understanding of student learning and development; and better instructional decisions for individual students and groups of students, which in turn lead to improved student learning outcomes.

In the absence of other information, a teacher at the start of a new school year would look to the students' results on the previous year's summative assessment to get a sense of the range of student achievement in the class, initial student groupings, and a general sense of where instruction may need to start. Research in the context of interim assessment (e.g., Goertz, Olah, & Riggan, 2009; Means, Padilla, DeBarger, & Bakia, 2009) has identified challenges with respect to what data are available to teachers, the timeliness of the data, organizational structures that are in place to support the use of the data, and tools to help teachers act on the data. The Winsight Assessment System will need to attend to these challenges.

The Winsight Assessment System is intended to help teachers qualitatively calibrate their classroom judgments and expectations of quality work using the interim assessment results, and understand student competency development better to make more informed appropriate instructional decisions. Teacher reflection is a highly valued practice, as seen in evaluation frameworks such as the Framework for Teaching (Danielson, 2013). In several countries, a process of social moderation is used to help teachers calibrate classroom judgments in order to use those data for accountability purposes (Wyatt-Smith, Klenowski, & Gunn, 2010). In social moderation, teachers often review student work examples and score them together in order to develop a common understanding of the expectations embodied in standards. The process of reviewing and scoring student work provides teachers with meaningful data on which to reflect. In the Winsight context, data to support both individual and joint teacher reflection will come from the interim testlet assessments. A similar approach was previously recommended by Morrison, Healy and Wylie (1995) in the context of using standardized Assessment Units in the UK as a mechanism to support teachers in calibrating their classroom judgments. Research will be needed to understand whether and how these data from standardized assessments, teacher discussion and reflection support calibration of teacher understanding of the standards and the expectations to which they hold students. In summary, there is empirical evidence of the value of teacher reflection and, from a logical perspective, using results from the interim testlets to spur such reflection regarding their classroom expectations is an important aspect of ensuring coherence across the system. It will be critical to both provide supports for teachers to engage in this practice, and to observe the impact of those supports.

The Winsight Assessment System is drawing heavily on the body of research from the Cognitively Based Assessment *of*, *for*, and *as* Learning (*CBAL*®) research initiative, which has been ongoing at ETS for almost a decade. This work has emphasized the disciplined development of LPs and KPs through reviews of empirical literature, cognitive lab studies and expert review (Graf & van Rijn, 2015) together with empirical validation of progressions (van Rijn et al., 2014). The LPs and KPs have informed assessment design for both summative and formative tasks (Bennett, 2010; Graf & van Rijn, 2015; Deane et al., 2015). Research suggests that teachers who have deeper content knowledge for teaching (CKT) are better able to facilitate learning based on the students'

understanding and needs (Ball, Thames, & Phelps, 2008; Baumert et al., 2010). Askew and Wiliam (1995) found that "Learning is more effective when common misconceptions are addressed, exposed, and discussed" (p. 8). A number of studies using treatment and control groups have demonstrated that when teachers use evidence of student learning to adapt instruction, student learning improves (Bergan, Sladeczek, Schwartz, & Smith, 1991; Fuchs, Fuchs, Hamlett, & Stecker, 1991). In summary, there is evidence to suggest that the kinds of insights into student learning that are captured by LPs and KPs can help teachers make sense of where students are in their learning and adapt instruction to meet students' needs for a positive impact on student learning.

Formative assessment practices include the use of clear learning goals and/or criteria for success, the elicitation of evidence of student understanding, the use of student self and peer assessment (discussed in the following section), and the provision of formative feedback. Unless teachers are explicit about the purpose of learning, students often do not understand the purpose of a specific lesson or how it fits within a larger sequence of learning (White & Frederiksen, 1998). Both quantitative and qualitative empirical research has found that students need to understand what they are learning and how they will be assessed to support one another effectively and develop a sense of autonomy (Tell, Bodone, & Addie, 2000). Research also has demonstrated the importance of teachers developing and/or selecting questions that provide quality evidence regarding student thinking and misconceptions (National Council of Teachers of Mathematics, 2000; Wylie & Ciofalo, 2009); systematically collecting evidence from all students in the classroom (Tobin, 1987), and deepening more engaging classroom discussions (Marshall & William, 2006). When teachers sustainably engage students in these practices, teachers are able to collect the evidence that allows them to better plan instruction based upon students' current understanding (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Graham, Milanowski, & Miller, 2012; Mevarech, 1983). In addition, empirical research studies have demonstrated that students need feedback that helps them both understand what the goals are, where their work is in relation to those goals, how to close the gap in order to move learning forward, and time and opportunities to act on that feedback (Black & Wiliam, 1998; Patthey-Chavez, Matsumura, & Valdes, 2004; Shute, 2008).

## Claim 4: Impact on students

> *When students engage with all components of the Winsight Assessment System, they have opportunities to grapple with important content in ways that model good teaching and learning. Winsight formative assessment resources encourage students to engage in reflective practices on their own learning and to provide feedback to peers using carefully designed structures. When students reflect on their learning and that of peers, students better understand what quality work looks like, take more ownership of their own learning and engage more with feedback, both in terms of providing feedback to peers and acting on feedback from peers and teachers. When students better understand what quality work looks like, take more ownership of their own learning and engage more with feedback, student learning improves.*

The claims associated with the purple boxes in the theory of action graphic focus on the student's role in formative assessment. Together, these claims form an argument that providing opportunities for students to reflect on their work and to both give and receive peer feedback will result in a range of positive intermediary effects that in turn ultimately impact student learning in a positive way.

Research has shown that peer feedback is effective when students are provided with explicit structures for providing feedback to each other, and routines for working collaboratively (King, 1992; Mercer, Dawes, Wegerif, & Sams, 2004). Research has also demonstrated that in a range of contexts frequent tests can improve student learning (Rohrer & Pashler, 2010; Hinze, Wiley, & Pellegrino, 2013; Paul, 2015) — for example, spaced quizzing with feedback improved student performance on subsequent assessments (McDaniel et al., 2013). A range of empirical studies have focused on a number of specific noncognitive outcomes that are a result of students engaging in self-assessment. These outcomes have included students developing internal attributions (Cohen, Raudenbush, & Ball, 2003), a feeling of empowerment (McDonald & Boud, 2003), improved self-efficacy (Andrade,

Wang, Du, & Akawi, 2009), and a sense of autonomy (Brookhart, Andolina, Zuza, & Furman, 2004). In addition to these noncognitive outcomes, studies have shown that when students engage in self-assessment and reflection, learning improves (Cohen, et al., 2003; White & Frederiksen, 1998).

In summary, to the extent that the Winsight formative assessment component and supporting resources provided to teachers can make learning goals explicit, provide tasks that are structured in ways that facilitate opportunities for students to both give and receive feedback, and encourage more student reflection on their own learning, we expect to see positive impacts on learning outcomes.

## References


Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, *102*(4), 287–302.

Askew, M., & Wiliam, D. (1995). *Recent research in mathematics education 5–16*. HM Stationery Office.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it so special? *Journal of Teacher Education, 59*(5), 389–407.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133–180.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement, 8*(2–3), 70–91.

Bergan, J. R., Sladeczek, I. E., Schwartz, R. D., & Smith, A. N. (1991). Effects of a measurement and planning system on kindergartners' cognitive development and educational programming. *American Educational Research Journal, 28*(3), 683–714.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics, 57*(2), 213–227.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4), 499–531.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119–142.

Danielson, C. (2013). *The framework for teaching: Evaluation instrument*. Princeton, NJ: Danielson Group.

Deane, P., Sabatini, J., Feng, G., Sparks, J., Song, Y., Fowles, M., & Foley, C. (2015). Key practices in the English Language Arts (ELA): Linking learning theory, assessment, and instruction. *ETS Research Report Series, 2015*(2), 1–29.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*(3), 617–641.

Goertz, M. E., Olah, L. N., & Riggan, M. (2009). Can interim assessments be used for instructional change? Policy Brief. RB-51. *Consortium for Policy Research in Education*.

Graf, E. A., & van Rijn, P. W. (2015). Recommendations based on observations from a mathematics assessment. *Handbook of Test Development*, 165.

Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and promoting inter-rater agreement of teacher and principal performance ratings. *Online Submission*.

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language, 69*, 151–164.

Honig, M. I., & Coburn, C. (2007). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy*.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*(4), 496–520.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, *27*(1), 111–126.

Marshall, B., & William, D. (2006). *English inside the black box: Assessment for learning in the English classroom*. Granada Learning.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27*(3), 360–372.

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice, 10*(2), 209–220.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision making in schools: Teacher access, supports and use*. Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Prepared by SRI International, Menlo Park, CA.

Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal, 30*(3), 359–377.

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*(4), 415–429.

Morrison, H. G., Healy, J., & Wylie, E. C. (1995). Teacher knows best: A solution to the marks-to-levels problem in National Curriculum Testing. *British Educational Research Journal, 21*, 175–182.

Munk, D. D., & Bursuck, W. D. (2001). What report card grades should do and communicate. *Remedial and Special Education, 22*(5), 280–287.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Patthey-Chavez, G. G., Matsumura, L. C., & Valdes, R. (2004). Investigating the process approach to writing instruction in urban middle schools. *Journal of Adolescent & Adult Literacy, 47*(6), 462–476.

Paul, A. M. (2015). Researchers find that frequent tests can boost learning. *Scientific American, 313*(2), 1–7.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*, 406–412.

Shute, V. J., Hansen, E. G., Underwood, J. S., & Razzouk, R. (2011). A review of the relationship between parental involvement and secondary school students' academic achievement. *Education Research International, 2011*, 1–10. Retrieved from: http://www.hindawi.com/journals/edu/2011/915326

Tell, C. A., Bodone, F. M., & Addie, K. L. (2000). *A framework of teacher knowledge and skills necessary in a standards-based system: Lessons from high school and university faculty*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research, 57*(1), 69–95.

Tompson, T., Benz, J., & Agiesta, J. (2013). *Parents' attitudes on the quality of education in the United States*. Retrieved from The Associated Press-NORC Center for Public Affairs Research website: http://www.apnorc.org/projects/Pages/parents-attitudes-on-the-quality-of-education-in-the-united-states.aspx

van Rijn, P. W., Graf, E. A., & Deane, P. (2014). Empirical recovery of argumentation learning progressions in scenario-based assessments of English Language Arts. *Psicología Educativa, 20*(2), 109–115.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and instruction, 16*(1), 3–118.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgment practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, policy & practice, 17*(1), 59–75.

Wylie, E. C. (2017). *Winsight assessment system: Preliminary theory of action*. (ETS Research Report No. RR-17-26). Princeton, NJ: Educational Testing Service.

Ysseldyke, J., Dennison, A., & Nelson, R. (2004). Large-scale assessment and accountability systems: Positive consequences for students with disabilities. Synthesis Report 51. *National Center on Educational Outcomes*.

Zapata-Rivera, D., Nabors-Olah, L., Vezzu, M., Biggers, K., Leusner, D., & Bertling, M. (2013). *Research and development on score reports for CBAL/CCSSO report system*. Princeton, NJ: Educational Testing Service.