

POLICY FORUM

Teacher Empowered Assessment: Assessment for the 21st Century

William Dardick and Jaehwa Choi
The George Washington University

There are several options to solve the problems surrounding the burden of standardized testing on students as it is often used link to teacher evaluations. One solution can be found through changing policies and practices to merge modern educational thinking with technological and psychometric advancements. This paper suggests the importance of the educational assessment system to transform from contemporary standardized testing that tends to burden students and teachers alike to an information communication technology based student assessment system with embedded valid arguments to evaluate a teacher's ability as an assessor of student ability. An overview into modern thinking on evidentiary reasoning and Information and Communication Technologies (ICT)-based assessment related to educational assessment is presented. Policy recommendation for a conceptual framework called Teacher Empowered Assessment (TEA) is introduced, based on modern assessment theory and advances in technology. Implications and future work of the TEA are also considered.

Keywords: assessment; standardized testing; validity argument

Lawmakers, teachers, parents, organizations and citizens regardless of being an advocate or opponent of the educational system demand a great deal from an educational assessment system. These demands and expectations take many forms. For example, lawmakers expect that no child is left behind which led to legislation (No Child Left Behind Act of 2001; U.S. Government Printing Office, n.d.) with general anticipation to monitor progress using large-scale tests. The standards of assessment are set with the expectation of states to demonstrate compliance and competency through testing. Parents expect assessments to inform us of their child's ability and needs. Teachers expect assessments to aid them in their judgment of student ability in a content area and for formative assessment to help diagnose a student's performance. School administrators expect assessment to assist them in many levels of decision making, including evaluation of our teachers. Test developers expect to maintain quality assessments to provide reliable and valid evidence of students' learnings on the underlying concepts and skills. All of these stakeholders expect our children to learn and expect assessments to be fair.

With so many expectations and desirable qualities of assessments, the existing system of standardized testing seems to overburden students and teachers (as well as administrators and parents). It should also be no surprise to the practitioners that researches, meetings, articles,

organizations, and websites dedicated to sorting out the pros and cons on the issue of standardized testing have been formed. For the interested reader many of the pro and con arguments for and against standardized testing can now be found on websites (e.g. ProCon.org). Arguments on both sides are inspired by empirical studies, logical and emotional arguments, public opinion, and range from issues in teaching and student burden to cost and cheating. Some of these arguments involve: comparable standards for students; time spent preparing; teaching to the test; teacher guidance in curriculum; reliable and valid measures of student ability; cheating by teachers and students to inflate scores; and accuracy in measuring ability. Standardized testing and the burden on students is an unresolved issue in today's world of education. The Washington Post recently reported on a study that shows the heavy workload of students, "Study says standardized testing is overwhelming nation's public schools" (Layton L. 2015). Everyone including the U.S. Department of Education has an action plan to reduce testing burden. In the Washington Post article they discuss President Obama's pledged to reduce the testing load on students. The President questions, "If our kids had more free time at school, what would you want them to do with it?" This sentiment represents the frustration of parents that students are overburdened with testing. The United States Department of Education's (DoED) "Fact Sheet: Testing Action Plan" (2015), states it is "... essential to ensure that tests are fair, are of high quality, take up the minimum necessary time, and reflect the expectation that students will be prepared for success in college and careers." They go on to state, "In too many schools, there is unnecessary testing and not enough clarity of purpose applied to the task of assessing students, consuming too much instructional time and creating undue stress for educators and students". The Fact sheet (DoED, 2015) lays out principles for fewer and smarter assessments and calls for smarter models of student assessment such as the Innovation Lab Network (Council of Chief State School Officers, 2016).

Coupled with the issue of assessment burden on students, many states link teacher's evaluation and promotion to student success on standardized tests. According to the American Statistics Association (ASA; 2014) statement on using value-added models for educational assessment, many school districts and states use Value-Added Models (VAMs) in the educational accountability systems. VAM is a catch all term for a variety of statistical methods that use student test scores and characteristics (at the student and school level) to estimate effectiveness of teachers and schools (Braun 2015). These are used increasingly as a "component in high-stakes decisions such as determining compensation, evaluating and ranking teachers, hiring or dismissing teachers, awarding tenure, and closing schools." There have been proposals of teacher accountability based on student growth data some of which have led to policies in some states. The use of this form of evaluation should not be considered without understanding issues of teacher training and student growth. In the ASA statement, these models are "generally based on standardized test scores, and do not directly measure potential teacher contributions toward other student outcomes." They further state "VAMs typically measure correlation, not causation" potentially leading to teacher rankings with unintended consequences. There is agreement that VAM scores are useful but have high variance and low stability over years (Braun 2015). There is a large amount of instability in VAM estimates due to methodological and design decisions (Darling-Hammond 2015; Goldhaber 2015). VAM should not be used in isolation as "errors are an inevitable part of any system of teacher classification" (Goldhaber 2015). Braun (2015) suggests more attention be paid to improving the validity of underlying tests, more time spent in implementation and to include the educational and political context in design of accountability systems.

Standardization increases accuracy and comparability of test scores while aiding in the interpretation of scores as defined by the test developer (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Standardized tests are designed with psychometric properties of validity, reliability and fairness. There is however a great deal of assessment conducted at the classroom level that could be combined with many of the desirable characteristics of standardized testing to reduce the student burden and help evaluate teacher performance. Changes should be made regarding the current standardized testing practices and general assessment of students, including issues related to student burden and teacher performance without maintaining the desirable qualities, usefulness and interpretability of test scores.

The current paradigm of assessment is nested in antiquated forms of technology used in testing systems and obsolete thinking about the form of assessment. The central issue with most standardized forms of assessment is that we are still using industrial revolution thinking from the earlier part of the 20th century. Standardized assessments are typically summative in nature and mass produced. The advantages of this model are the desired qualities expected from education testing, such as evidence for validity, reliability and fairness. Unfortunately they are often dropped into the student's life artificially with disruption to learning, sometimes misaligned to the classroom needs. Further, in the classroom setting we have formative and summative assessments developed by content experts (e.g. teachers) which provide a more natural and less disruptive setting however these assessments are void of the rigorous psychometric evidence found in standardized assessments. Teachers are typically not trained in large scale standardized test development nor do they have the time to build assessments on their own with the desirable properties of standardized assessment. The standardized test is trusted to be psychometrically reliable and valid for primarily summative purposes, while the merit of classroom assessment is aligned to specific curriculum, less- intrusive and used for both formative and summative purposes.

There are several options to solve the problems surrounding the burden of student performance and teacher evaluations. For example, two extreme and opposing options might be to leave things as they currently exist or drop everything. Leaving things as they exist suggests that the ends outweigh the means, i.e., the need for summative evaluation of students and teachers is necessary for the education system and it is the best we will have for the foreseeable future regardless of the burden. Another and opposite solution might be to completely drop both standardized assessments and teacher evaluation. This would completely eliminate the burden but we would know very little if anything regarding comparability of students or teachers. There are many other solutions that lay somewhere in between. Adapting and retaining standardized tests outside the classroom is possible while lowering some of the burden on students. This would not resolve the full issue of the testing burden or teacher evaluation but would be an improvement and is among the directions as described by the DoED (2015). An alternative for teacher evaluation, VAM with improved teacher measures, could be developed to consider performance, likely at a very high cost. What is desirable is a solution that can ease the testing burden, keep students in the classroom, provide some insight into teacher evaluation, and empowering teachers for assessment.

In this paper we argue that the problem of burdensome classroom assessment and standardized testing on students coupled with the desire to have evidence of student outcomes in the evaluation of teachers is found in today's technology merged with a new way of thinking

about assessment systems. The introduction provides some of the foundation of the argument that all stakeholders (with a focus on students) are burdened by the current model and teachers evaluation linked to outcomes may not measure potential teacher contributions. In the following sections of this paper we will: (1) provide some background into modern thinking in evidentiary reasoning related to educational assessment and technologies based assessment; (2) offer a conceptualization of Teacher Empowered Assessment (TEA) based on modern thinking and advancements in assessment; (3) present implications of the TEA for policy; (4) suggest policy recommendations; and (5) make concluding remarks. The authors contend that a solution to the problems with assessment can be found through changing policies to merge modern educational thinking with technological advancements. Modernizing student assessment and teacher evaluation will improve the educational system as a whole.

BACKGROUND OF MODERN THINKING IN ASSESSMENT THEORY AND PRACTICE

Evidentiary Arguments

The 2014 standards of educational and psychological testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) provides desirable criteria that can be incorporated into the quality of any test or assessment. However, the practical considerations of the classroom assessment do not make it feasible to conduct formal validation. We still desire tests that are valid, reliable, and fair to our students. Assessments used to make high stakes judgment of student performance can impact outcomes beyond the grade assigned in a classroom are of particular interest. The standards have progressed toward modern evidentiary-based thinking for support of validity, reliability, and fairness.

The use of validity as arguments and the use of evidence as inference in educational assessment for the development of quality assessments have been around since before the turn of the century (Embretson, 1983; Cronbach, 1988; Messick, 1989, 1994; Kane, 1992, 2006, 2013; Mislevy, 1994, 2006; Mislevy, R., Steinberg, L., & Almond, R. 2003: sources are provided both as reference and additional reading). One useful framework for validity arguments and evidence-based reasoning is the Toulmin framework (1958), which can be used in assessment arguments to support a claim based on data where a warrant is backed by evidence to justify the claim (Kane 2006, 2013). The use of mathematical probability in modeling arguments of evidentiary reasoning helps guide claims and traces backwards toward the rationale (Mislevy 1994). Arguments are never perfect and can of course become substantially complex but can be evaluated and empirically supported to help us understand the link between many educational assessments and claims regarding students. Using evidence and validity argument, we can do much more than evaluate a student using the traditional multiple choice item. We can provide a rule-based system of various forms to measure complex tasks even incorporating human raters as needed.

Evidence-centered assessment design (ECD) is a method used to design and deliver educational assessments using evidentiary arguments (Mislevy, Almond & Lukas, 2003). ECD originated at Educational Testing Service (ETS) with Robert J. Mislevy, Linda S. Steinberg, and Russell G. Almond (1999) and merged advances including technology with evidentiary

reasoning to develop inference from task in more complex settings. In a special issue of *Applied Measurement In Education*, (Huff and Plake, 2010) four articles with application from the College Board's Advanced Placement Program® (AP®) detail how ECD “can provide many benefits to large-scale test development, resulting in a clear, coherent and comprehensive evidentiary argument for the valid interpretation of scores”. Thinking through the chain of reasoning in an ECD based assessment system is that reasoning and advances in technology can serve also as the foundation for improvement in assessment, teaching, and learning.

Information and Communication Technologies (ICT)-based assessment

The advances in today's technology permit student assessment to be conducted anywhere, on different types of devices and nested in a conceptual design framework. ICT-based assessments coupled with current trend in computer based testing and Auto Item Generation (AIG; Choi, 2016) can provide us with the latest technological advancements linked to assessment. It would be a useful policy to move toward the common technical devices such as smartphones, laptops, watches or whatever the device of the future may be; implemented in teaching, learning and assessment. When we consider the cutting edge of technology today it may be automated cars wirelessly communicating to each; television content providers tracking our patterns to predict our favorite show; buying everything with the wave of a phone; and a seamless TEA nested in the learning environment and capable of formative and summative assessment at a moment's notice. The suggested use of ICT is not hope for useful technology and advancements in the future, but to use what is readily available today at our fingertips, with an eye toward advancement.

ICT-based assessment tools can enable teachers to monitor student performance more frequently both inside and outside of the classroom setting. Teachers can provide students with individually tailored and instantaneous feedback, to determine next steps for classroom instruction. Through the integration of advanced ICT-based technology, future classroom practices will be grounded in a deeper understanding of how students think and learn within the context of a content domain. Assessment can be seen as directly supporting the cognitive process of reasoning from evidence through a variety of sources. Teachers can utilize this information to re-design or adapt assessments through easy to use ICT-based application(s) to provide students with more detailed and individualized feedback about particular qualities of their progress and what they can do to enhance learning.

Using ICT-based assessment tools it is possible to collect evidences regarding students' understanding and learning in various ways — by responding to a teachers' questions in a classroom setting using an application such as a tablet PC or phone, working with a computerized web-based assignment application at home, or even playing a game-based learning application. All of these ICT server-client applications can work harmoniously to provide and update such assessment information regardless of students' time zone and location. Additional development of novel, targeted task formats could be developed and measured with creative, innovative technologies but for brevity those above are well established already and useable today.

Technology and tools of modern assessment including the use of a multifaceted validity argument can be used to redesign our assessment and evaluation system in education. The five ECD language and layers (Mislevy, 2013) of Domain Analysis, Domain Modeling, Conceptual

Assessment Framework (CAF), Assessment Implementation, and Assessment Delivery are applicable. For example, Domain Analysis helps us think through the task in the classroom and evaluation considerations of the teacher while Domain Modeling helps to build assessment arguments both for students and teachers. The producers and ideas of ECD are the foundation of a Teacher Empowered Assessment (TEA) here presented and we will mainly focus on those elements that are not explicate to ECD. The interested reader might consider further reading on ECD or validity arguments which is beyond the scope of the current paper. (Kane, 1992, 2006, 2013; Messick, 1989; Mislevy, Steinberg, & Almond 1999, 1994; Mislevy, 1994, 2006; Mislevy, R., Steinberg, L., & Almond, R. 2003)

TEACHER EMPOWERED ASSESSMENT

Teacher Empowered Assessment (TEA) is a conceptual framework grounded in the modern thinking of validity and evidence based arguments together with the advancements of ICT and psychometrics. The goal of TEA is to provide a solution that folds together standardized and classroom formative and summative assessment of students with teacher evaluation concerns to form a new way of assessing and assisting students' learning. TEA is a proposed student and teacher centered assessment to provide accurate and comparable judgements of student ability in educational curriculum, estimate teacher's skill in judging student ability and use of assessment tools, and support psychometric instruments with teacher judgement. TEA examines student ability first grounded in common anchor items or systematic tasks that have been well established and the evidence for validity and reliability amongst a common pool of items/tasks are well known. AIG could be used for its advantages but it is also acceptable to have complex tasks or traditional items embedded in the TEA as well. These items/tasks would have psychometric rigger as the standardized tests used today, however the method of delivery and setting would be very different, as well as the intended use. The selected standardized items/tasks would be used in the same way a subset of anchor items are used for comparability amongst two or more equivalent assessment.

Teachers are permitted to select amongst these established items and other alternative items. These alternative items might be some combination of self-developed items/tasks for curriculum proposes or selected from preexisting set of items to support judgments made about students. Essentially, items or tasks could be standardized, but could also be developed by the teacher to make formative assessment and summative decisions regarding a student. Multiple forms of assessment can be used to provide feedback, evidence for student ability and make final weighted decisions regarding students. This idea is not very novel although implementation on a large scale through traditional means would not be feasible. Teacher's selection of existing psychometric evaluated items/tasks along with the use of their own developed or selected items/tasks can be linked together.

The established anchor items/tasks can help further provide evidence of a teacher's ability to use assessment tools to make judgments about students. This analysis of teacher ability can provide evidence of a teacher's ability to use and or deign assessment tasks so as to empower them to design overall assessments of students. As with any sound evidentiary argument this inference will be incorporated into the model from the very beginning so claims regarding teachers follow from their choices, not just from student outcomes. The foundation of the teacher claim is the argument that highly effective teachers can make sensible judgements about the

students in their classroom. With more reliable and valid teachers judgement, comes improved assessment and instruction of students. Teachers' ability surrounding assessment tools would be one set of measures in teacher evaluation but not the only measure. The complexity of teacher evaluation should give rise to numerous well-reasoned arguments that make alternative inference for evaluation regarding teacher. This judgement measure is our current focus given its ability to be readily tied in to the assessment system. Formative feedback to the teacher regarding their judgments of students can help teachers utilize assessments to support student learning and target their own ability to evaluate students as an independent rater of student ability in the classroom. Finally when used as a whole, the TEA will be able to provide evidence for teacher ability to design and use assessment instruments, suggest estimates of teachers' ability around judgments they make regarding student ability and use this information to make weighted judgments about student ability. One method to produce student ability or scores would be to use judgements of teachers, weighted by a reliability measure of teacher ability (i.e., a human rater reliability score) alongside any standardized items and tasks.

This multidirectional assessment system for Teacher Empowered Assessment (TEA) transforms assessment to be student centered and vested through the teacher for evaluation of both student formative and summative assessment as well as used to provide feedback to the teacher on their judgments through reporting. When coupled with advances in technology, this assessment incorporates the desirable qualities of standardized testing with the strength of formative assessments and summative teacher judgments. It eliminates the need for outside the classroom assessment and supports a teacher in the difficult task of developing, designing and using tests. Through this assessment a precision about the teacher's ability as an expert judge and evaluator of student performance becomes more reliable. For example, teacher ability as a rater of students can be estimated along with a precision or variance around that score. The teacher should be able to make reliable and valid judgements about students with multiple teaching tools/items/tasks. Teacher ability and judgement can be demonstrated when a proportion or subset of the tools/items/tasks for assessment used are independently created or selected by the teacher, while another subset are standardized tools/items/tasks.

Over time as more information about that teacher is gathered, the score becomes more precise and it would be known with more accurate judgements of teachers regarding students. The teacher rater or judgement score could eventually impact the proportion or subset of independent assessments versus those that are standardized. This will increase teacher ability to provide quality assessments, anchored with highly scrutinized core items and tasks while providing evidence for valid, reliable and fair judgments of students, with formative feedback on assessment in the classroom, empowered by teachers. There would be no need to go outside of the classroom to make judgments about students or track their progress.

The TEA can serve as a platform where stakeholders can work collaboratively using assessment information. In an education system, teachers, administrators, and policy makers can use the assessment system information as a shared knowledge base system for monitoring and understanding how students learn material and what characteristics of proficiency are important to measure. All tasks can be performed through an ICT network system on the on-going growth data. The new assessment's tools, resources, and materials are able to perform a modern scientific analysis on how students learn. This new assessment framework can produce the basis for redesigning the roles of classroom and large-scale assessments, the curriculum and instruction, so that all the TEA's components work toward a more coherent set of learning goals than in the traditional assessment paradigm.

It is imperative to note that these new paradigms for assessment can signify a substantial departure from the conventional and commonly used types of assessments in current practice. A substantial level of improvements are required for such assessments to produce useful (i.e., reliable and valid) assessment information. For example, the Computer Adaptive Formative Assessment (CAFA: Choi, Kim, & Yoon, 2012) system is currently available for the math domain and more development is needed to provide robust formative feedback on students' progress both digital (Choi, Kim, & Yoon, 2014; Choi, Kang, Kim, Dardick, & Zhang, 2015) and paper (Choi, Kim, & Yoon, 2016) workbook formats.

For the digital format workbook, an empirical research experiment was conducted in South Korea at the high school level as a pilot-test of the workbook and its features. Students and teachers participating in the focus group appreciated the system-generated problems with the same pattern but with unique equations, which enabled students to better understand and master specific standard-aligned items (Kim et al., 2014). In a second empirical study (Kim, H., Kim, I., & Jung, Y., 2015), students provided feedback on the use of the workbook application. The results show that 87.7% of the participants are positive about the convenience of the experience with the workbook. When surveyed on the convenience of the system, 91.2% indicated that the experience was positive. General satisfaction with the system is indicated by 91.1% of parents. The survey results indicate that the most desirable quality in the tool is the function, which generates additional similar items for incorrectly answered items. This feature allows users to obtain additional practice.

The formative assessment system developed in this study can be used to customize learning effectively by providing users with instant feedback. This system can determine if the user (student) is proficient in each standard, and will also determine where the student may need additional instruction or extra practice. Kim et al. (2014) argued the system could contribute to improving public education as a means of providing customized academic support in the standards (such as CCSS). The CAFA system can be seen as one of the gateways to developing TEA. Full implementation of CAFA system requires work and research on how to design and develop assessments not only for theoretical aspects (e.g., the underlying properties of measuring learning), but also for technical and systematic aspects of assessment (e.g., architecture of the server-client assessment platform in ICT environments).

THE TEA IMPLICATIONS FOR POLICY

Computer and communications technologies are making it possible to implement more complex assessments like TEA in practice. Student learning can be measured with a fine level of detail, across multi-media simulations of real-world contexts, and in ways that are highly integrated with classroom instruction practices. Examples of complex simulation-based assessment utilizing ECD from engineering and medicine are presented by Robert Mislevy (2013). Technology is making formative assessment more feasible even in complex and potentially dangerous settings. It would not be a stretch to consider a classroom as an environment to demonstrate how students can receive on-the-fly individualized feedback as they work under a system powered by teachers, TEA. Feedback would now be individualized, instantaneous, and with possibly even more detail than what a classroom teacher alone could have provided in classroom.

There is urgency for policy to enable the rise of a “genuine 21st century version of

assessment tool” as demonstrated by the testing burden on students coupled with the evaluation and burden of teachers and assessment. Many challenges in educational practice can be overcome by synergistically combining educational assessment and technological advances. As emphasized in the Common Core State Standards (CCSS: Common Core State Standards Initiative 2015), it is important to secure tools and resources that can facilitate high levels of student achievement. Applications, such as the TEA, will enhance student learning and academic achievement in schools by making assessment and curriculum more accessible, feasible, and clear to students, teachers, parents, and other education stakeholders. This is because the full scope of student achievement, growth and learning is mapped by virtue of Information and Communication Technologies (ICT)-based assessment tool. Further there is a great deal of assessment transparency for all involved.

Utilizing a centralized computer system connecting assessment applications; teachers in the TEA carry less burden of assessment data management. They can now spend more time on using information to adapt instruction for the class as whole and for the individual student on the basis of personalized understanding and learning progress. In this assessment framework, teachers have vastly upgraded understanding and richer picture of students’ learning progress, by domain, at a group or classroom level, and for the individual student. Furthermore, teachers can better observe patterns of learning progress over time, e.g., typical learning pathways or typical learning bottlenecks. Monitoring feedback on achieving learning goals (e.g. short-term, mid-term, and long-term) can also be shared instantly and/or regularly with students and parents as a part of the instruction. Because of individualized and instantaneous feedback provided by the TEA, students can be actively engaged in online or offline activities, from home or in the classroom. The resulting efficiency and accessibility will enable more time for peer and self-assessment in order to implement the criteria for high-quality work and develop metacognitive skills emphasized in the CCSS or other initiatives.

The information from the TEA is not limited to the formative purposes. Stakeholders can also use information for a summative purpose to obtain current reflection and feedback regarding a student’s overall learning progress while providing progress report to others. These summative assessments can also benefit from key advantages of the ICT-based assessment system, i.e. instantaneous, individualized feedback and growth. Upper level accountability systems primarily focused on summative assessments, such as state tests, can more efficiently monitor and strengthen the same educational goals and principles about learning that are operating at the classroom level. Large-scale assessments will be able to individually set valuable learning objectives for each student to achieve. In this paradigm, such large-scale assessments can broadly access the desired outcomes for learning by using a variety of client applications centralized in an assessment server system (e.g., CAFA platform: Choi, Kim, & Yoon, 2012). For example, it is possible to efficiently perform a large-scale assessment for the on-the-fly homework assessment information combined with assessment data on classroom work produced during the course of instruction.

In light of new capabilities of a TEA, all stakeholders in education, e.g., educational policy makers, educators, and the public can expect more than the conventional usages (i.e., ranking and comparing) of large scale assessment data in summative assessment portraying current test results. First, performance on large-scale assessments can be instantaneously delivered via ICT tools so that students, parents, and teachers can also use such information immediately. Due to its formative nature, the assessment framework can provide more assessment information on the concepts and processes entailed at different levels of competence.

For example, such assessments will be able to provide information on how a competent student performs on a Common Core standard in math, compared with another student who is in different level of proficiency. This is possible because such centralized assessments possess large-scale (big-picture) assessments and can provide different kinds of strategies in problem solving, interpretation, approaches and procedures, explanations, and results that distinguish among various levels of proficiency.

POLICY RECOMMENDATIONS FOR TEA

Our recommendations are for an investment in a framework like TEA that utilizes the flexible framework of ECD, validity arguments for both students and teachers through the currently available advancements of ICT and psychometric to support all forms of assessment and provide education opportunities. ECD models of assessment in simulation (Mislevy 2013), College Board's Advanced Placement Program® (AP®) and the CAFA platform (Choi, Kim, & Yoon, 2012) are already in use and among those platforms that have elements of a future assessment and evaluation system for education. The authors recommend setting up a policy that a school system could consider all forms of formative, summative, standardized assessment and teacher evaluation as a linked model to be considered as one complex assessment and evaluation system. We also recommend a centralized computer system connecting assessment so student can do homework on a phone or laptop, take a test on computer or tablet and acquire a grade on a classroom activity that can report psychometric properties. We further recommend teacher judgment be a weighted consideration in comparability of students, dependent on the evaluation of teacher ability.

Working together the school system, teachers, administrator, law makers and test developers can build the system that suites the need of the system while meeting and legal requirements. Technology supported by AIG and teacher judgement should be incorporated into the assessment system (e.g. standardized testing, classroom and homework) to decrease student burden and improve the educational system.

CONCLUSION

The proposed TEA moves educational assessment into the 21st century. Future research in this assessment framework will include both a technical and detailed cognitive blueprint in ECD language surrounding teacher and student inference. The full potential of a new assessment significantly depends on educational practices and adoption of new policies. Some constraints and regulations should be relaxed or changed to realize the full benefits of a merger between the ICT and assessment. The new assessment described in this article does not necessarily follow the current practice of computer adaptive testing or computer based testing rule that, i.e., all students must take on-demand or paper-and-pencil assessment under strictly uniform circumstances. Instead it is rooted in the philosophy of ECD practices, advances in technology, and arguments about teacher quality. Appreciating the full benefits of new forms of assessment clearly depends on making harmonious alterations in curriculum and instruction.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA
- American Statistical Association (2014). ASA Statement on Using Value-Added Models for Educational Assessment. Retrieved from http://www.amstat.org/policy/pdfs/asa_vam_statement.pdf
- Braun, H. (2015). The value in value added depends on the ecology [Special issue]. *Educational Researcher*, 44, 127-131.
- College Board (2016). How AP develops courses and exams. Retrieved from <https://advancesinap.collegeboard.org/ap-course-exam-development>
- Council of Chief State School Officers (2016). Retrieved from http://www.ccsso.org/What_We_Do/Innovation_Lab_Network.html
- Choi, J. (2016). *Automatic Item Generation – Next Generation Item and Test Development*. Assessment, Testing, and Measurement Technical Report Series. The George Washington University, Washington DC.
- Choi, J., Kang, M., Kim, N., Dardick, W., & Zhang, X. (2015). A smart way of coping with Common Core challenges - Introduction to CAFA SmartWorkbook. *Journal of Educational Issues*, 1(2), 70-89. doi:<http://dx.doi.org/10.5296/jei.vli2.8381>
- Choi, J., Kim, S., & Yoon, K. (2012). *CAFA (v 1.0 Beta) System: Computer Adaptive Formative Assessment System for Educational Services [Computer System]*. CAFA Lab, Inc.
- Choi, J., Kim, S., & Yoon, K. (2014). *CAFA SmartWorkbook (v 1.0 Beta) System: Computer Adaptive Formative Assessment Client Application for Common Core Math Workbook [Computer System]*. CAFA Lab, Inc.
- Choi, J., Kim, S., & Yoon, K. (2016). *K-Math Workbook Grade 6*. Clarksville, MD: CAFA Lab.
- Common Core State Standards Initiative (October 24, 2015) Retrieved from <http://www.corestandards.org/>
- Cronbach, L. (1988). Five perspectives on validity argument. In *Test Validity*, H. Wainer and H. Braun, Eds. Lawrence Erlbaum, Hillsdale, NJ, 3-17.
- Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? [Special issue]. *Educational Researcher*, 44, 117-126.
- Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce [Special issue]. *Educational Researcher*, 44, 87-95.
- Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527-535.
- Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed.), 17-64. Westport, CT: American Council on Education and Praeger.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1, 1-73.
- Kim, H., Kim, I., & Jung, Y. (2015). Introducing a formative assessment system to support customized education (II) – Development of an Online Item Management System. *Korea Institute of Educational Curriculum and Evaluation, Research Report. RRE 2015-10*.
- Kim, H., Park, J., Jung, Y., Park, S., Kim, C., Lee, C., & Choi, J. (2014). Introducing a formative assessment system for supporting individually tailored education (I) – Blueprint of on and off-line formative assessment system. *Korea Institute of Educational Curriculum and Evaluation, Research Report. RRE 2014-9*.
- Layton L. (2015, October 24) Study says standardized testing is overwhelming nation's public schools. *The Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/study-says-standardized-testing-is-overwhelming-nations-public-schools/2015/10/24/8a22092c-79ae-11e5-a958-d889faf561dc_story.html
- Messick, S. (1989). Validity. In *Educational Measurement*, 3rd ed., R. Linn, Ed., American Council on Education, Washington, D.C., 13-103.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Mislevy, R. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59, 439- 483.
- Mislevy, R. (2006). Cognitive psychology and educational assessment. In R. Brennan (Ed.), *Educational*

- measurement* (4th ed.), 257–305. Westport, CT: American Council on Education and Praeger.
- Mislevy, R. J., (2013) Evidence-centered design for simulation-based assessment. *Military medicine* 178(10S):107-114
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A Brief Introduction to Evidence-Centered Design (CSE Report 632). CA: Center for Research on Evaluation, Standards, and Student Testing.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999, December 4). Evidence-centered assessment design. Retrieved from http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html
- Mislevy, R., Steinberg, L., & Almond, R. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-66.
- U.S. Department of Education (October 24, 2015). Fact sheet: Testing action plan. Retrieved from <http://www.ed.gov/news/press-releases/fact-sheet-testing-action-plan>
- U.S. Government Printing Office (n.d.). H.R.1-No Child Left Behind Act of 2001: 107th Congress (2001-2002). Retrieved from <https://www.gpo.gov/fdsys/pkg/PLAW107publ110/html/PLAW-107publ110.htm>