MAXIMIZING THE FUNCTIONALITY OF THE FLORIDA ITEM BANK AND TEST PLATFORM FOR SCIENCE ASSESSMENT

by

Sally Childs Sanders

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2013

STATEMENT OF PERMISSION TO USE

In presenting this professional paper in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the MSSE program shall make it available to borrowers under rules of the program.

Sally Childs Sanders

July 2013

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

ABSTRACT

This study investigated Florida science educators' perceptions of the functionality and value of the state Item Bank and Test Platform. Evidence from educator surveys, along with anecdotal evidence from meeting notes, shows a shift in educators' perceptions from little to no familiarity with the Item Bank and Test Platform to a favorable perception of its value as an accountability tool for the Next Generation Sunshine State Standards. The successful use and sustainability of the Item Bank and Test Platform as a valid and reliable measure of student achievement depends on maximizing its functionality and value to teachers.

INTRODUCTION AND BACKGROUND

Project Background and Rationale

Background and History

The US Department of Education awarded a Race to the Top Grant to Florida in August 2010. A key component of the award focused on the standards and assessment assurance area, including the development of high-quality, balanced assessments. This assessment initiative is managed by the Office of Race to the Top Assessments (ORTTTA) at the Florida Department of Education (FDOE). The mission of the ORTTTA is to develop and implement a statewide assessment item bank, the Florida Item Bank and Test Platform (IBTP). The FDOE has contracted with a testing company to manage the development of the IBTP, which will serve as an online repository for assessment questions (or items) and as an online delivery system for assessments.

The IBTP is a four-year project that will be operational in June 2014. The bank will include assessment items for English language arts, mathematics, science, social studies, and Spanish. Science items are being developed for all benchmarks in grades K–8, as well as all benchmarks in four high school science courses: chemistry, biology, physics, and Earth and space science. This is a very ambitious project; the projected total number of science items is over 11,000. The purpose of this item bank is to provide teachers, schools, and districts with high-quality, standards-based test items that have been written, reviewed, and approved by teachers in Florida. One rationale for this initiative is to provide test items designed to assess Florida's Next Generation Sunshine State Standards (NGSSS), instead of relying on test items from outside sources such as

publishers' test banks or other vendor-generated test items, which may be aligned to the National Science Education Standards, but not to any state's science standards. The IBTP project will provide a bank of valid and reliable assessment items that specifically target Florida's state standards for science, the Next Generation Sunshine State Standards.

One unique aspect of this project is that many of the items will be written and reviewed by science teachers. Some science teachers in Florida have had opportunities to review items for the Florida Comprehensive Assessment Test (FCAT) and End of Course (EOC) assessments, but this is their first opportunity to actually create science assessment items for a state item bank.

Each item goes through several levels of review for quality assurance. The IBTP will have five levels of functionality, described in Table 1, when it is fully operational in June 2014. Teachers can select items and generate a customized quiz or test (Level 4); schools can build site-level tests (Level 3); school districts can generate district-level tests (Level 2); and the state can create tests for school improvement and accountability purposes (Level 1). Level 5 will serve as a repository for non-secure sample items and will be available to the public.

Table 1
Functionality Levels of the Florida IBTP

Access Levels	User Access					Purpose
	FDOE	School Districts	Schools	Teachers	Public	
1	✓					Accountability tests for low performing schools
2	✓	✓				District-level tests
3	✓	✓	✓			School-level tests
4	✓	✓	✓	✓		Classroom-level tests
5	✓	✓	✓	✓	✓	Sample items

The IBTP is designed to meet a variety of assessment purposes, including benchmark, formative, interim, and summative assessments**.** The IBTP provides a diagnostic tool for science teachers to identify areas of weakness in curriculum and instruction and address those weaknesses. Because this tool allows teachers to adjust curriculum and instruction before students take an accountability test, it will be beneficial to both teachers and students. Each school district in Florida designs its own science curriculum based on the state science standards identified in the state course descriptions. The IBTP will provide a valid measure of any district's curriculum, as both the curriculum and the items in the bank are driven by the state science standards and state course descriptions.

The IBTP will have the functionality for searching by benchmark (from the state science standards), grade level, course, depth of knowledge level, item type, and keyword. The IBTP will include science items of seven different types, as shown in Table 2.

Table 2
Item Types for the Florida IBTP

Item Type	Description	Point Value
Selected Response	Students select from provided options (Multiple Choice, Matching, True-False)	1
Gridded Response	Students enter a numerical response and bubble the corresponding digits on a grid	1
Short Response	Students provide a one or two word response	1
Constructed Response	Students provide a brief written response	2
Extended Response	Students provide an extended written response	4
Essay	Students write an essay response	6
Performance Task	Students perform a task	1-10

In my role as Science Assessment Specialist with the FDOE ORTTTA, my primary job responsibility is working with the contractor to manage the development and implementation of the science items. This means that I have a professional interest in ensuring that the item bank is made of high-quality, valid, and reliable test items, and that the test platform is user-friendly for Florida science educators. At the end of the grant period, we will have approximately 11,000 standards-based science items in the bank and an operational test platform; however, because federal funding for the project ends in 2014, we are looking at ways to continue the functionality beyond the contract period. The main planned source of continuity is the ongoing contribution of additional items into the bank by teachers in order to sustain the IBTP. The realization of this project depends on the extent to which teachers use and value the IBTP.

Rationale and Purpose

The idea for my action research project came from the IBTP initiative. The IBTP will serve as a repository for test items, test generator, and online test delivery system, and may potentially be used not only for measuring student achievement, but also for

evaluating teacher effectiveness. Many school districts are choosing to include student achievement measures as one tool for teacher evaluation instruments. The Florida Comprehensive Assessment Test (FCAT) provides a measure of student achievement in science for grades 5 and 8 and an end-of-course assessment in Biology 1. The IBTP will fill in the gaps for assessing student achievement in other grades and courses.

My research is intended to evaluate science teachers' perceptions of the IBTP in order to maximize its utility and value. The outcomes of this research will be used to guide the ORTTTA in the continuing development and implementation of the IBTP in order to ensure maximum functionality and sustainability of the system, and will also help me to serve Florida's science teachers and students by providing effective guidance and support for the use of the IBTP. The goals of my project are to explore and analyze educators' perceptions of the IBTP in order to establish a baseline status for the IBTP, and to use the knowledge gained to make process improvements and content improvements to the IBTP.

Research Questions

With the goal of exploring educators' views of the IBTP, I selected this question as my primary focus question*: How can the functionality and value of the Florida Item Bank and Test Platform be maximized?* This question is relevant for me professionally because a large component of my job is working to ensure the success of the IBTP and providing assessment support to science teachers. The primary goal of the ORTTTA is for teachers and students in Florida to have a rich source of high-quality, standards-based test items for measuring student achievement and learning gains throughout the year.

Having access to this item bank will also be a time-saver for teachers, as it will provide a direct source of items developed specifically to measure student proficiency on the NGSSS. Teachers using the IBTP will not have to spend time creating new items or generating tests from other sources. Instead, they will have the capacity to quickly generate standards-based quizzes and tests, and when they choose to create new items, these items can be added to the item pool for use by other teachers.

The IBTP will also serve as a form of professional development at two levels. The science teacher cohort serving as item writers and reviewers is receiving training on creating and reviewing assessment items for science and they are receiving support from the testing vendor and the Office of Race to the Top Assessments. All teachers in Florida will have the opportunity and capability to use the system to create items, access items, and generate customized tests. They also will have access to the training materials and will receive feedback on the items they create.

One sub-question related to this project is: *What are science educators' perceptions of the Florida Item Bank and Test Platform?* This question seems important for establishing a baseline status of teachers' views of the IBTP. If teachers do not have a positive perception of the value of the IBTP, they will not be likely to use it. A second sub-question is: *How can science educators' perceptions be used to improve the functionality and value of the Florida Item Bank and Test Platform?* This question is important for driving continuous improvement and sustainability of the IBTP.

Timeline

The IBTP timeline began with the hiring of content and assessment specialists for the ORTTTA and the selection of an assessment vendor through a Request for Proposals (RFP) process in 2011. The ORTTTA staff has participated in Professional Learning Community (PLC) workshops since 2011 in order to build skills in assessment development. Development of item specifications and items took place in 2012–13. The statewide implementation of the IBTP is planned for the 2013–14 school year.

The first five item batches have been developed and uploaded into the IBTP. Each batch consists of 600 – 1200 science items. The overall timeline for item writing and review for these first five batches is August 2012 – June 2013. Table 3 provides the timeline for input and review of the first two batches of assessment items; the development of these item batches is concurrent with the time period for this research project.

Table 3
Timeline for Batches 1 and 2 for the IBTP

Batch	Date	Action
1	8/1/12 – 9/24/12	Items developed
	9/24/12 – 10/16/12	Items reviewed by science educators and other science experts
2	8/20/12 – 11/1/12	Items developed
	11/6/12 – 1/6/13	Items reviewed by science educators and other science experts

Once the items for a particular batch have been developed and submitted into the IBTP system, the period of time for item review spans approximately four to eight weeks.

Collegial Support

I am fortunate that the nature of my project provides for built-in collaboration. I work with a team of dedicated content area assessment specialists in the ORTTTA. The team's primary task is developing and implementing the IBTP. The team consists of one assessment specialist each for English language arts, mathematics, science, social studies, and Spanish. Each team member brings strengths and expertise useful for meeting the goals of the IBTP project. Collectively, the ORTTTA team provides skills in editing and proofreading documents, assessment development, curriculum development and support, database applications, and data analysis.

My colleagues have helped throughout my action research project by providing feedback and guidance. Because they are also important stakeholders in the IBTP project, they were able to offer valuable insight into the data collection and the analysis of the data. They are also directly involved in using the results of the data to improve the IBTP. Our team holds weekly meetings, and a portion of that time is used for problem solving and generating process improvement ideas.

This action research project has the potential to impact many teachers and students throughout the state, as well as district assessment and curriculum coordinators. After identifying relevant questions to pursue for this project, next steps included conducting a literature search to learn more about current research and findings related to science assessment.

CONCEPTUAL FRAMEWORK

Concepts and Theories Related to Assessment

The focus of my action research project is looking at the effectiveness of the Florida Item Bank and Test Platform (IBTP). This is a timely topic in Florida, as in other states, because of the high-profile nature of high-stakes testing. One of the driving forces behind the development of the IBTP was a new Florida statute that required each school district to provide assessments for each course offered in the district. Another requirement of this statute is a new plan that has been put into place for teacher evaluation in Florida, which may include student performance on assessments as a factor in evaluating teacher performance. Although this plan is mandated by state statute, it does not identify specific components of the plan. A review of related literature helped to identify previous research related to perceptions of large-scale assessments, and also helped to guide the direction of my research.

Research about the effectiveness and value of high-states testing has shown that parents' perceptions are mostly favorable (Osburn, Stegman, Suitt, and Ritter, 2004). The parents of 5th grade students were surveyed, along with the teachers and students. The 19 survey questions were grouped into three constructs: parental involvement and interest in testing, testing climate, and overall stress and anxiety. A large majority of parents reported a high value and interest in testing for the first construct. Most parents reported a positive perception of testing climate for the second construct, as well as a perception of moderate levels of test anxiety and stress for their children. The results of this small-scale survey do not indicate negative views of testing on the part of parents.

Another study that was relevant to my research targeted about 700 teachers in the state of Florida (Jones and Egley, 2004). The results indicate that most of the teachers surveyed had a negative perception of the FCAT (Florida Comprehensive Assessment Test) and its impact on curriculum, teaching and learning, and student and teacher motivation. The strongest negative response (93.7%) indicated that teachers do not think that school grades should be based on students' FCAT scores. The FCAT assesses student achievement in only three subjects – mathematics, English language arts, and science. So, the strong views against using FCAT scores for school grades may be based on the idea that it's unfair to make a judgment about a school's success based on just those three subjects, without also considering other subjects and other aspects of education. The positive effects cited by the teachers were related to the role of testing in accountability. The results and implications of this study are very relevant for my work in a state education agency related to implementing curriculum and assessment policy, especially for the appropriate use of testing and its role in teacher evaluation programs and school grading policy (both of which are in place in Florida). The IBTP and the FCAT target the same state science standards and both may be used for accountability purposes. The study was conducted in the spring of 2002 after four years of FCAT administration. This study categorized teachers' perceptions into the following themes:

- use and accuracy of the test
- effects on curriculum
- effects on teaching and learning
- effects on student and teacher motivation

For each of the themes, teachers reported more negative than positive perceptions, but the study also clearly shows that teachers are in favor of accountability; they simply disagree with much of the policy related to the use of the FCAT. The study also looked at research in other states which also indicates negative perceptions of testing on the part of teachers. The most useful aspect of this study for my research is the abundance of data, especially qualitative survey responses, and the potential for using this data directly in my job to implement the new testing program. This particular study was especially helpful in guiding the direction and details of my data collection techniques.

Some researchers propose a philosophical view that high-stakes testing could potentially lead to a "default" philosophy of education based on valuing a small subset of knowledge and skills rather than the full range of skills we associate with student achievement (Gunzenhauser, 2003). Gunzenhauser (2003) describes this default philosophy as a negative perspective, in that testing for accountability purposes drives curriculum and limits instructional innovation and creativity. One particularly striking point in this article is the contrast between the viewpoint of psychometricians that tests are fallible and imperfect and the general viewpoint of the education community that sees tests as absolute measures of student achievement. Tests can be used to provide our best judgment of a "true score" and should not be considered as an absolute measure of student achievement, but an approximation. An awareness of this apparent contrast between psychometricians' views and those of the larger education community was useful for my research because this conflict is sometimes observed in the education community. This concept is important in the development and implementation of policy

related to testing because it identifies some common misconceptions which can impact the use of evidence from testing.

Two research methods that are relevant to my research are descriptive statistics and qualitative interpretation (Hsu, 2005). Both of these made sense for my study because I have collected data to describe existing conditions (teachers' perceptions of the state IBTP) and to interpret teachers' survey responses in a way that will lead to effective guidance from the state education agency to help teachers successfully implement the IBTP. This study looked at the relative use of various data collection strategies, but did not go into detail about specific procedures within each strategy.

Many education leaders are calling for a renewed focus on assessment as evidence for educational decision-making, and moving away from the limited view of assessment as high-stakes testing (Reeves, 2007). This new direction of assessment focuses on the fundamental purposes of assessment: improvement of teaching and learning and leadership decision-making. He suggests that educators must move away from a bell curve model, at least in terms of a central framework for assessment. In his view, it is more important to compare student performance to an objective standard than to the performance of other students. This fits with the trend towards criterion-referenced assessment and away from norm-referenced assessment. Similarly to states' No Child Left Behind assessments, the items in the IBTP are intended to measure individual student mastery of Florida's Next Generation Sunshine State Standards, not to compare an individual student's performance with other students' performance.

Assessment should be viewed as an integral and essential part of teaching and learning that provides information to help students move closer to achieving the learning targets (Guskey, 2007). Guskey (2007) describes the importance and appropriate use of formative assessment. He reminds us that the end goal is for students to achieve the learning targets, and assessment is a valuable tool to meet that goal. His description of formative assessment mirrors the goals of the IBTP project. We are building the IBTP in hopes that teachers will value and use it as a resource for improving teaching and learning, not as another testing burden.

Other researchers have built the case for professional learning communities as a forum for teachers to help each other become better at effectively using formative assessment (Wiliam, Lee, Harrison, and Black, 2004). The article cites studies supporting the use of formative assessment as the single most important factor in teaching practice for increasing student achievement, compared to other interventions such as increasing teachers' content knowledge and reducing class size.

The research conducted by Guskey (2007) and Wiliam, et al (2004) was especially relevant to my study because these authors describe formative assessment as assessment *for* learning, as opposed to high-stakes testing, which is described as assessment *of* learning. Their research was helpful and inspiring in terms of reinforcing my belief that assessment can and should be one of the most powerful tools for improving instruction and student achievement.

This review of related literature helped me to understand the current research on perceptions of assessment, specifically related to teaching and learning, and helped me to

narrow the focus of my research to the potential impact of the IBTP on teachers' practices and effectiveness and students' achievement, and how to improve the functionality and value of the IBTP. The literature review showed evidence of mixed views of assessment on the part of teachers and parents, but was encouraging because of the trend towards an acceptance of the value of formative assessment as a tool for improving instruction and student achievement.

METHODOLOGY

Descriptive Study

This project is a descriptive study which does not include the use of a typical intervention or treatment. The study is based on a research question, and two sub-questions, intended to help me understand an existing situation (science teachers' current perceptions of the IBTP) and how this information can be used to maximize the functionality and value of the IBTP.

To collect data for addressing my research questions, I used three techniques, summarized in Table 4 below. The basis for each of the data collection instruments is to get a sense of teachers' views of the IBTP in order to maximize the positive impact of the IBTP on curriculum and instruction, student motivation and learning, and teacher motivation and effectiveness. The first two data collection instruments consisted of surveys administered to groups of science educators. The meeting minutes from the ORTTTA served as the third data collection instrument.

Data Collection Instruments

Data Collection Instrument #1 – The Item Writer and Item Reviewer Surveys

The Florida Department of Education sent an announcement to each of the state's school districts in the summer of 2012, inviting nominations for item writers and item reviewers. This provided a unique opportunity for science educators to have more direct involvement in creating assessment items. The content specialists in the Office of Race to the Top Assessments reviewed the applications and selected the first cohort of item writers and reviewers. Participants were selected to represent the current demographics of

Florida based on the following criteria: geographic distribution, small and large districts, rural and urban districts, ethnicity, and gender. The participants in the study consisted of 39 science item writers and 72 science item reviewers, and represented the diversity of the state.

This first cohort of item writers and item reviewers for the IBTP participated in training via WebEx in August and September of 2012. The WebEx training was offered on five different occasions (evenings and weekends) to accommodate the teachers' schedules. The WebEx training and the survey were designed by the testing contractor, with guidance and input from the ORTTTA, specifically for the IBTP. The item writers began creating items within a week of completing the training, and the item reviewers began reviewing items soon after the first items were submitted. Each item writer and reviewer participated in a survey immediately after completing the training. These surveys were designed to elicit feedback from the trainees in order to identify potential problems and misconceptions related to the IBTP. The results from these surveys provided a source of data for addressing the action research questions.

This first cohort of science educators provided a unique perspective as they were essentially "piloting" the new item writing process and review tool developed for the IBTP project. Both the item writer training and the item reviewer training focused on characteristics of high quality assessment items and related concepts such as difficulty and cognitive complexity of items and principles of universal design. Because the survey was customized for participants who had already received the training, there was no opportunity to provide a pilot survey before collecting data from the participants.

Both the Item Writer Survey (Appendix A) and the Item Reviewer Survey (Appendix B) consisted of 14 questions to elicit feedback. The final survey question provided an open-ended question and field for recording a response. Feedback from the item writers and reviewers was used to provide information about the functionality of the IBTP.

Data Collection Instrument #2 – The Science Educator Survey

In December 2012, I presented a session at the annual state Science, Technology, Engineering, and Mathematics (STEM) Conference. The topic of the presentation was *Florida's Interim Assessment Item Bank and Test Platform Project*. At the end of the presentation, session participants were invited to complete a survey to capture their views of the IBTP. There were 37 survey participants, consisting of science teachers and district science coordinators from across the state of Florida. The survey was designed to elicit the participants' initial perceptions of the IBTP. Their feedback serves as a representative sampling of educators' current perceptions of the IBTP, bearing in mind that these educators chose to participate in the state STEM conference and may have more familiarity with the IBTP than many other educators. A survey was used to gather information because time constraints at the conference would not allow for individual interviews with the session participants. The Science Educator Survey (Appendix C) asked questions about science educators' knowledge of the IBTP, as well as their opinions of its value, in order to understand their level of familiarity with the IBTP and their perceptions of how it will be used.

The survey administered at the STEM conference provided a different data set than the Item Writer Survey and Item Reviewer Survey, which provided information specifically about the usefulness and quality of the item writer training and the item reviewer training.

Data Collection Instrument #3 – Meeting Minutes

The Content Team of the ORTTTA conducts weekly meetings to discuss issues and topics related to the IBTP. These meetings provide a forum for the team to brainstorm and to share strategies for successful implementation of the project. The meeting minutes assisted with a more accurate reflection on the perceived value of the IBTP and helped identify ways to improve the project. The minutes were captured during each meeting in a document and then stored in a network folder.

One section of the meetings is reserved for communication issues. The ORTTTA team has a communications plan that provides strategies for disseminating information about the IBTP to educators and for providing support to educators for use of the IBTP. Meeting minutes in the comments and communications sections from August 2012 through March 2013 were sorted and analyzed.

Validity and reliability are two assessment components that the ORTTTA team must constantly adhere to. It is also important that the data collection tools for this project be valid and reliable measures of educators' views of the IBTP. For each of the data collection instruments, internal consistency reliability of the tool is determined from several items measuring different aspects of the same concept. For example, the Item Writers and Item Reviewers Surveys included clusters of items in three different

concepts, and the Science Educators Survey included several items addressing different aspects of the concept of usefulness and value of the IBTP. Reliability of the data from the meeting minutes came from multiple anecdotal notes communicating similar results. The content validity of the data collection instruments came from evaluations of the survey questions and the meeting notes by members of the ORTTTA team in order to ensure that the surveys were asking the right questions and did not include extraneous questions.

The combined results from these data sources are being used to guide the sustainability plan for the IBTP. Table 4 shows how the three instruments were used.

Table 4
Data Collection Matrix

Research Questions	Data Collection Instruments		
	1: Item Writer Survey and Item Reviewer Survey	2: Science Educator Survey	3: Minutes from Content Team Meetings
How can the functionality and value of the Florida Item Bank and Test Platform be maximized?	August – September 2012	December 2012	August 2012 – March 2013
What are science teachers' perceptions of the Florida Interim Assessment Item Bank and Test Platform?	August – September 2012	December 2012	August 2012 – March 2013
How can science teachers' perceptions be used to improve the functionality and value of the Item Bank and Test Platform?	August – September 2012	December 2012	August 2012 – March 2013

The research methodology for this project received an exemption by Montana State University's Institutional Review Board and compliance for working with human subjects was maintained. The exemption letter is provided in Appendix D.

DATA AND ANALYSIS

Process

The Science Educator Survey, Item Writer Survey, and Item Reviewer Survey each consisted of closed-ended items, as well as open-ended items. Although the responses to the survey are qualitative, the data is represented quantitatively by tallying the number of responses for each closed-ended item and entering the data into an Excel spreadsheet. For both the closed-ended and open-ended items, the thematic analysis system was used. This analysis consisted of studying the data for patterns to help code the data, analyzing the data and codes to identify patterns and themes, developing answers to the research questions based on the identified patterns and themes, and providing interpretation in narrative form. In looking at both types of data, the final interpretation was conducted through the lens of the research questions.

Data Analysis

Data Analysis for Instrument #1– The Item Writer and Item Reviewer Surveys

Quantitative score reporting was used to analyze both the Item Writer Survey results and the Item Reviewer Survey results (Appendix A and Appendix B.) The questions were the same across the two surveys, with the only difference being the focus on item writing or item reviewing. The first ten questions on both surveys were related to characteristics of high-quality items and the mechanics of item construction. The next three questions (#11–13) were related to the quality of the training. The final question (#14) was related to overall effectiveness of the training.

For all questions on both surveys, the overall scores were greater than 3 on a 4 point scale (Appendix E and Appendix F.) Analysis of the three themes addressed in the survey (mechanics of item writing, quality of training, and overall effectiveness of training) showed the item reviewer group responded more favorably to each of the themes. Table 5 shows a summary of the survey results for both item writers and item reviewers.

Table 5
Summary of Florida IBTP Item Writer Survey and Item Reviewer Survey Results (*N*=111)

Item Writer Survey and Item Reviewer Survey – Summary of Results			
	Mechanics of Item Writing	Quality of Training	Overall Effectiveness of Training
Average Score for Item Writers	3.44	3.67	3.44
Average Score for Item Reviewers	3.54	3.68	3.68

The average rating for the set of questions about item writing (items 1–10) was 3.44 for the item writer group and 3.54 for the item reviewer group. Overall, the item reviewers rated this part of the training slightly higher than the item writers did. This may reflect a higher comfort level for reviewing items than for creating items. Item writers may feel a greater responsibility and a greater need for thoroughly understanding the mechanics of item writing. The next set of questions (items 11 – 12) was related to quality of the training. The two groups had similar ratings (3.67 for item writers and 3.68 for item reviewers.) For overall effectiveness of the training in terms of preparing the participants to serve as item writers or item reviewers (item 14), the item reviewer group gave the training a score of 3.68, compared to the score of 3.44 for the item writer group.

This may be due to the different levels of responsibility for item writers and item reviewers. The lower ratings for training quality by the item writers may indicate their need for more thorough training in writing valid and reliable items. One strategy to increase item writers' confidence would be to have them practice writing some science items and conduct a follow-up session to review these items together, with feedback from the trainers.

It is encouraging that the survey responses provided by this cohort of science item writers and reviewers indicated a favorable view of the item writer/reviewer training. Their responses also indicated a probable need for additional training or support during the item writing/reviewing process. Because this cohort of science teachers took the initiative to submit an application to serve as item writers/reviewers, their perceptions of the IBTP are likely to be more favorable than the general perceptions of science teachers throughout the state. It is anticipated that as this cohort completes their item writing and reviewing assignments, they will share positive feedback with other science teachers so that others will be motivated to use the IBTP and volunteer to serve as item writers and reviewers. One strategy for process improvement would be to invite this cohort of science item writers and reviewers to serve as mentors to subsequent groups of item writers and reviewers for the IBTP project.

The final question on both surveys invited participants to provide comments. The survey results included 32 entries in this field. The comments were sorted into the three categories represented in Table 6 below. These categories were identified based on the patterns and topics addressed in the teachers' comments and suggestions.

Table 6
Summary of Qualitative Feedback from Item Writers and Reviewers (N=32)

Feedback Categories	Number of favorable comments	Number of unfavorable comments
1. Mechanics of item writing/reviewing	15	2
2. Quality of training	5	6
3. Technology and system issues	1	3

The high percentage of favorable comments related to the mechanics of item writing/reviewing may indicate that the science teachers who received training were receptive to learning how to write high quality assessment items, and that they may have gained useful information from the training. One item writer commented, "This is very comprehensive. I've been on FCAT review committees and I thought I knew how to write test questions but I learned more specifics in this training." The unfavorable comments in this category were related to specific information that was not provided in the training. For example, one teacher responded, "Provide direction on writing scoring rubrics for the open ended questions." Another teacher commented, "You mentioned item specifications but you have not provided them to us. Will we have the item specs available to us for reviewing items?" These comments exposed unintended gaps in the training to be addressed in a follow-up WebEx for this cohort and also with the next groups of item writers/reviewers.

The feedback for the first category reflected favorably on the content of the training; however, the feedback for the second category was more evenly distributed

between favorable and unfavorable comments related to the actual delivery of the training. One teacher commented, "The training manual you provided us does not match up with the PowerPoint. They should follow the same sequence." Another teacher responded, "The training was too heavy on details like how to add graphics to a question." This feedback indicated a need to address certain aspects of the training, such as ensuring alignment among the components of the training materials and adjusting the amount of time spent on each topic, based on its importance. There may also be a need to include opportunities for participant interaction.

The high percentage of unfavorable comments related to the third category indicated a strong need to address technical issues with the online item authoring system. One comment related to technical issues was, "The system is very slow and difficult to navigate." Another teacher commented, "I was not able to get the online calculator tool to work." These responses may indicate technical issues in the system. It is helpful for the ORTTTA team to reflect on this data to monitor system improvements.

This qualitative feedback from item writers and reviewers has helped identify ways to improve the content and quality of the item writer/reviewer training, as well as target technical issues to monitor throughout the item development process. As the ORTTTA team spends more time using the online item development system, the team is better able to monitor the issues and to provide support to the item writers and reviewers.

Data Analysis for Instrument #2– The Science Educator Survey

The Science Educator Survey was administered to 37 science educators attending the researcher's presentation at the annual state STEM conference in December 2012.

This group of educators consisted of district science coordinators, secondary science teachers, and elementary teachers with an interest in STEM education. The questions on this survey were designed to elicit participants' thoughts on the IBTP and its usefulness for measuring science proficiency.

Question 1 was used to correlate overall views with district size, using a demographic table from the Florida Department of Education Office of Accountability, Research, and Measurement, that categorizes school districts by population size. Question 2 was used to correlate overall views with educator experience. Question 3 was used to correlate overall views with science teachers' involvement as IBTP item writers and reviewers in their districts. Table 7 shows a summary of the data from questions 1-3.

Table 7
Summary of Responses to Questions 1-3 from Science Educator Survey (N=37)

1. District Size	Small 22%	Medium 62%	Large 16%	
2. Years of Science Education Experience	<6 30%	6 – 15 43%	>15 27%	
3. District has Science Teachers Serving as Item Writers or Item Reviewers for IBTP	Writers and Reviewers 22%	Writers 16%	Reviewers 35%	Neither 27%

The majority of survey participants identified their district size as medium. Seventy-three percent (73%) of participants had 15 or fewer years of science teaching experience. For the third section of data regarding district participation in item writing and item reviewing, it is important to note that this data is for the 37 attendees at the STEM presentation, so it does not represent all districts. Some districts with science item

writers and reviewers were not represented at the STEM session. For future data collection, it will be important to administer surveys to these districts, in order to have representation from all districts involved in item writing and reviewing. Another factor to consider for this data set is that some of the session participants had no knowledge of whether their district had science teachers serving as item writers or item reviewers, and were thus unable to accurately report these numbers.

Items 4-8 were designed with a range of 1-3, with 1 representing a negative response and 3 representing a positive response. The goal of this survey was to provide an overall snapshot of educators' views (positive, neutral, or negative) of the IBTP immediately following a presentation on that topic.

Quantitative score reporting was used to evaluate the data for questions 4-8 on the Science Educator Survey. Table 8 shows a summary of the science educators' responses.

Table 8
Summary of Responses to Questions 4-8 from Science Educator Survey (N=37)

	Strongly Agree	Somewhat Agree	Disagree
4. My district is planning to provide professional development related to the Item Bank.	5%	19%	76%
5. My district will use the Item Bank for science assessment.	17%	78%	5%
6. The Item Bank will be a valuable and useful tool for meeting my district's needs for standards-based testing.	27%	70%	3%
7. Other science teachers in my district view the Item Bank as a valuable and useful tool for science assessment.	19%	76%	5%
8. My district is interested in contributing items to the Item Bank.	55%	40%	5%

While questions 1-3 were used to collect demographic data, questions 4-8 were used to establish current perceptions and expectations for use of the IBTP. Comparing the demographic data with the data on perceptions of the IBTP, there was no correlation between either district population size (question 1) and perceptions of the IBTP or the participants' years of educational experience (question 2) and perceptions of the IBTP. However, there was a positive correlation between current district participation in item writing and/or reviewing (question 3) and perceptions of the IBTP (questions 4-8). For example, 73% of the respondents indicated that their district currently has science teachers serving as either item writers, item reviewers, or both. The majority of these respondents selected the 'strongly agree' category for questions 4-8. This indicates stronger support for the IBTP in districts where science teachers participate as item writers and reviewers. This feedback is encouraging for the ORTTTA in terms of

continued recruitment of item writers and reviewers to increase involvement across the state, and consequently, increase the utility and value of the IBTP for more teachers. It would also be interesting to collect data to determine if the item writers and reviewers are actively using the system for test generation once it becomes operational.

Question 4 was about participants' district's plans to provide professional development for use of the IBTP. Seventy-six percent (76%) of the educators disagreed with this statement. This may be due to timing; at this time, districts may not have plans in place for providing professional development related to the IBTP.

Question 5 was about their district's plans to use the IBTP. This question scored a majority of responses in the mid-level range, with the remaining responses in the high range. This is encouraging, as it indicates an interest in using the item bank. One participant commented, "My district is small and we do not have the same resources as larger districts. We need this item bank and we need to work in partnerships with other districts to make our tests." This feedback is helpful because the ORTTTA can facilitate partnerships among districts for building common assessments.

For questions 6-7 about the respondents' perceptions of the value and utility of the IBTP and their estimate of other teachers' perceptions, the majority of participants indicated a mid-level ranking, with almost all other responses in the high category. This may be due to educators' familiarity with other measures of student achievement, such as the FCAT, and the potential for these measures to improve teaching and learning.

Question 8, about sustainability of the IBTP, was the only question that scored a majority of responses in the high category. This may mean that educators are interested in having a direct impact on the quality of the IBTP by contributing items.

Questions 9-10 were open-ended items intended to elicit specific needs and questions. Comments for these questions were mainly focused on asking the FDOE to provide professional development on the use of the IBTP. One participant commented "We are having assessment workshops next summer. Can your office send someone to work with us?" Another participant asked "We heard that districts will get a chance to use the item bank in the spring of 2013; when will we have access?" These questions reinforce an important objective of the ORTTTA team – to provide assessment support to school districts. We are planning multiple avenues to disseminate information about the IBTP to all districts in Florida. We will participate in district assessment workshops, create tutorials on use of the IBTP, post information on the website of the FDOE, and release statewide announcements through other venues, such as Florida's NSTA Science Matters and state organizations for science teachers and district science supervisors.

Data Analysis for Instrument #3– Meeting Notes

The meeting minutes from the ORTTTA Content Team meetings provided another source of data. The ORTTTA team of assessment specialists holds weekly meetings to discuss issues related to the IBTP project. One component of the ORTTTA communication plan calls for each of the content specialists to disseminate information to our colleagues throughout the state (teachers and curriculum specialists for our specific content areas.) The ORTTTA does this by presenting sessions at state and regional

conferences and by responding to questions from educators via email or phone. For the past year, the ORTTTA has provided communication updates at the weekly content meetings and shared anecdotal data from our interactions with educators in the classroom, as well as district curriculum specialists.

A review of the meeting notes from August 2012 through March 2013 shows that ORTTTA communications with stakeholders fall into two categories: educators' perceptions (and misconceptions) about the design and intent of the IBTP and educators' perceptions about how the IBTP will be used. For example, communication updates from October 2012, following the statewide meetings of curriculum specialists, indicated a misconception that the IBTP project will result in state-developed end-of-course tests that will be mandated for all public schools in Florida. The communication update from the October 30, 2012 meeting included the comment, "The science curriculum supervisors asked if we are building end-of-course exams for high school science courses, and they questioned why we would build another biology exam when we already have a state end-of-course biology exam." This type of comment may indicate that some educators have the wrong idea about the purpose of the IBTP. The intent of the IBTP is not for the FDOE to build tests, but instead to provide a bank of high-quality test items to meet the needs of school districts for building their own tests. Biology is one of the high school science courses addressed by the IBTP project, but the intent is to provide an abundance of assessment items for interim and formative assessment of biology students, not to provide another end-of-course test. Each opportunity that I have to make a

presentation to science educators, or to communicate one-on-one, is another chance to convey this message in order to clear up this misconception.

Another widespread concern among stakeholders was that the tests built from the IBTP will be used for teacher evaluation. The communication update from the December 11, 2012 meeting included the comment, "Some educators at the state STEM conference expressed concerns that tests built from the item bank will be used to evaluate teachers based on their students' performance on the test." Even though there is currently no state mandate requiring school districts to use tests built from the IBTP as either a measure of student proficiency on the state standards or as an evaluation of teacher performance, there is a perception among educators that this is the case.

Other team members reported similar concerns from their meetings with educators and curriculum supervisors for their content areas. The mathematics specialist on the ORTTTA team reported, "Math teachers and specialists are confused about the purpose of the IBTP. They think it is simply practice for PARCC." The Partnership for Assessment of Readiness for College and Careers (PARCC) is a 22-state consortium developing assessments based on the Common Core State Standards for Mathematics and English Language Arts. Florida is a member of this consortium and is serving as the fiscal agent. The communication reports from the fall of 2012 indicated a high level of concern and misconceptions about the IBTP and its intended use and function. Knowing the specific misconceptions allows us to create presentations intended to directly target these views and to answer questions teachers may have about the purpose of the IBTP.

A review of more recent meeting notes shows a trend from these negative perceptions (or sometimes lack of knowledge) of the IBTP reported in 2012 to more frequent communications and a more favorable view. The report for March 5, 2013 includes the comment, "ORTTTA content specialists are receiving requests from teachers for information about how they can be involved in the project and questions about when the item bank will be ready for use." If the cohort of item writers and reviewers is sharing positive feedback with their colleagues, this may account for the reports of more positive comments and questions.

Our team has taken advantage of many opportunities to communicate with educators in order to correct the misconceptions and to encourage educators to participate, both as item writers and reviewers, and as end users of the IBTP.

Based on anecdotal evidence from ORTTTA meeting minutes, the views of Florida educators have evolved from a more negative perception of the IBTP in the summer of 2012 to a more positive view in the spring of 2013 as more educators become involved in the project.

INTERPRETATION AND CONCLUSION

Changing Perceptions

This study provided information about the level of knowledge Florida science educators have about the IBTP, their perceptions of its functionality and value, and how their knowledge and perceptions have changed over the past year.

At the beginning of the IBTP project, science educators' views of the project may have been based on what they heard from their colleagues, and this may have resulted in some of the misconceptions about the project. In response to the first research question about science educators' perceptions of the Florida Item Bank and Test Platform, the results of this study provided evidence that many educators across the state initially had either a lack of knowledge or unfavorable views and expectations of the IBTP. This conclusion is based on data from the Science Educator Survey in December 2012 (Data Collection Instrument #2), which indicated that most participants did not expect their districts to provide professional development. Anecdotal evidence from the meeting notes of the ORTTTA (Data Collection Instrument #3) also indicated misconceptions and unfavorable views of the IBTP in August – September 2012.

The data from the item writer and reviewer surveys (Data Collection Instrument #1) shows that teachers who received training to become item writers and reviewers had mostly positive perceptions of the IBTP. Science item writers reported an overall score of 3.44, out of 4, for overall effectiveness of ITBP training and science item reviewers reported an overall score of 3.68. The qualitative survey responses indicated that the ORTTTA should be more thorough in planning the training for item writers and

reviewers. Two gaps identified by survey respondents were the item specifications and the need for specific training on how to write scoring rubrics for open-ended items.

Evidence from a survey of science educators attending the state STEM conference shows a range of familiarity and understanding of the IBTP. Educators from districts that have current IBTP item writers and/or item reviewers reported higher ratings for value and functionality of the IBTP. Evidence from this survey also indicated that, of the science educators sampled by the survey, most (97%) had at least somewhat favorable views of the IBTP. However, most of these science educators (76%) reported that their districts have not made plans to provide support to their teachers for use of the IBTP. Support for use of the IBTP was overwhelmingly, and incorrectly, identified as a responsibility of the FDOE. This conclusion was based on responses to the open-ended questions 9-10 from the Science Educator Survey (Data Collection Instrument #2). Of the 26 responses to these questions, 92% addressed their district's need for professional development provided by the FDOE ORTTTA.

Anecdotal evidence from the communications reports of the ORTTTA team of content specialists indicated a shift from either an initial lack of knowledge or negative perception to a more favorable perception among educators over the past year. This may be due, in part, to the efforts of the ORTTTA team to provide information and support at every opportunity when there is a captive audience of teachers, such as state curriculum conferences. Another factor in this change in perception may be the influence of the 39 science item writers and 72 science item reviewers currently working on the project. The

item writers and reviewers are classroom teachers and district science specialists who are likely to communicate their views of the IBTP to their colleagues.

Analysis of the evidence showed educators' perceptions evolving from negative and misinformed views in the fall of 2012 to more positive views and expectations in the spring of 2013. Anecdotal evidence from the ORTTTA meeting notes showed many misconceptions and unfavorable views in October 2012, following interactions of the ORTTTA content specialists with our colleagues at state content curriculum meetings. Meeting notes from March 2013 showed a shift to more informed views and more favorable views, based on interactions with our colleagues.

Evidence from the Science Educator Survey in December 2012 showed mixed results; participants indicated receptive views of the IBTP, but negative views of their district's capacity to support training and use of the IBTP. This may mean that the educators are open to the IBTP, but skeptical of their district's capacity to provide training. They may perceive this as a duty of the state Department of Education. This conclusion was based on the analysis of questions 4-8 from the Science Educator Survey. Each of these questions, related to views of the value of the IBTP, received a favorability rating (combining the responses for "strongly agree" and "somewhat agree") of at least 95%. The analysis of negative views of districts' capacity to support the IBTP was based on the favorability rating of 24%, based on responses to question 4 from the Science Educator Survey.

Another indicator of more favorable views of the IBTP was the mostly positive feedback from science item writers and reviewers. Science item writers rated the overall

quality of the training at 3.44 out of 4.00; science item reviewers rated the overall quality of the training at 3.68 out of 4.00. This may be due to participants' feelings of success in understanding the use of the IBTP following the training.

Using Data to Improve Value and Sustainability

In response to the second research question about how educators' perceptions can be used to improve the functionality and value of the IBTP, this study provided discrete, specific issues to target for improvement.

One aspect for improvement was specific components of the item writer/reviewer training and technology issues related to use of the online system for item development. This conclusion comes from analysis of the responses to the Item Writer Survey and Item Reviewer Survey, which indicated specific gaps in the training.

Another need for the ORTTTA to target is the need for educator support and guidance for the use of the IBTP. Analysis of the open-ended responses to the Science Educator Survey showed that the ORTTTA will need to build the capacity to provide support to districts for their assessment workshops. The ORTTTA is planning to provide access to the IBTP for a group of educators from each district in the state in summer 2013. The intent of this trial access is for educators to become familiar with the platform and to see a sampling of the items developed so far. The ORTTTA will elicit feedback from these educators in an effort to continuously improve the system and the quality of the items.

Other evidence from this study showed a need for continued efforts to communicate accurate information about the intended use of the IBTP. This conclusion

was based on the shift from misconceptions and unreceptive views to more favorable views reported in the meeting minutes of the ORTTTA team. When teachers are empowered to create and review assessment items for a statewide item bank, as opposed to all of the items being contributed by a testing company, they take on a sense of ownership and they want the IBTP to be successful because they have a hand in creating it.

Another important aspect to target is the inclusion of more science teachers as item writers and reviewers. Based on analysis of the Science Educator Survey, districts that currently have science teachers involved as item writers and/or reviewers reported more favorable views of the IBTP. Districts that have provided support and invested resources in the IBTP project had a higher degree of buy-in and were more pro-active stakeholders.

Evaluating this evidence has led to the conclusion that the ORTTTA can have a positive impact on educators' views of the IBTP, and that steps can be identified to ensure the success and sustainability of the IBTP.

VALUE

Benefit to Florida Students

The IBTP initiative has potential significance for all students in Florida. For example, science tests generated using the IBTP may be used not just for interim or formative assessments, but also as end-of-course tests to determine whether a student passes the course. The IBTP is not a mandatory testing requirement for school districts. In order for the IBTP to become a valuable tool for assessing student achievement, it must be embraced and valued by teachers. The results of this study can be used to improve teacher acceptance and use of the IBTP, which will have a positive impact on students by providing them with high-quality standards-based items for measuring achievement.

Benefit to Florida Teachers

The IBTP initiative has potential significance for all teachers in Florida, independent from the impact on their students. Teachers' job performance may be evaluated based on their students' performance on these tests. The IBTP also provides a source of items for formative assessment, so that teachers can use results to adjust instruction as needed. Science teachers who have an appreciation for the value of the IBTP, and who make effective use of the IBTP, will have a valid and reliable accountability tool for measuring their students' achievement of Florida's Next Generation Sunshine State Standards (NGSSS), and they can be confident that the results generated by this tool are valid and reliable. Based on the data that showed positive feedback from science item writers and reviewers, the ORTTTA is planning to offer

IBTP training to as many science educators as possible. The ORTTTA is also developing online modules that will provide information on the use of the system for generating tests and how to analyze student performance data. One strategy to foster continued use of the IBTP is to encourage science teachers who have used the system during the first year of operation to present a session at the annual conference of the Florida Association of Science Teachers.

Professional Benefit

This study has provided the researcher with an understanding of educators' views and expectations of the IBTP and offers evidence that these views are changing in a positive way. The study also provides ideas for strategies to continuously improve the functionality and value of the IBTP, with an emphasis on continuing to provide training and professional development on the use of the IBTP. It will be important to develop high-quality tutorials for use of the IBTP and to invite feedback from users of the system. The results of this study have indicated that thoughtful communication and support in the form of training workshops, presentations, and individual guidance can be effective in increasing educators' knowledge of the IBTP, and in gaining their support for standards-based assessment.

The ORTTTA will use this evidence in planning statewide professional development for educators and in offering support for districts to develop and use high-quality assessments.

The data from the surveys has identified misconceptions among science educators about the IBTP. One strategy I plan to implement, based on this data, will be modifying

the content and delivery of my presentations about the IBTP in order to target possible misconceptions. For example, I will directly address the need for including biology as one of the targeted courses for item development, even though Florida has a state end-of-course biology exam. The ORTTTA wanted biology teachers to have access to a rich source of high-quality, standards-based items covering the scope of the course so that their students' achievement of the standards can be measured throughout the year, not just at the end of the course. Of course, this statement is also true for all science courses in the IBTP project, but the misconception about biology arose from the fact that a state biology exam is already in place.

Analysis of the data identified gaps in item writer/reviewer training. The ORTTTA is making changes to the content of the training workshops in order to correct these gaps.

Data from this study showed that even though perceptions of the IBTP have become more positive overall, negative views and expectations still exist, especially related to professional development on the use of the IBTP.

One strategy I plan to implement to address this is taking advantage of every opportunity to communicate with science educators, individually and in groups, to address their questions and needs. It will be important to communicate the potential value of the IBTP as a resource for formative, interim, and summative assessments. This communication will include posts on the website of the FDOE, statewide announcements through venues such as the state NSTA Science Matters eblast, and sessions at the conferences of the Florida Association of Science Teachers (FAST), Florida Association

of Science Supervisors (FASS), and the state STEM conference. Building support through the FASS will be especially important, as this is a strong, very active group of dedicated science supervisors who provide support to science teachers in their respective districts. Having both organizations, FAST and FASS, as partners in the development of high-quality science assessment items for statewide use will be a powerful tool.

Another important strategy for addressing the professional development concern is the development of online modules that science educators can access 'on-demand'. The ORTTTA team is working with the assessment contractor to develop and implement these modules. Eliciting feedback from science teachers after they have used the system is another strategy for continuous improvement of the IBTP.

The ORTTTA has a challenge in the upcoming 2013-14 school year, which is the final year IBTP project. The challenge is to continue improving the quality of the item bank and its reception and use by Florida educators so that the ORTTTA will have support from educators throughout the state for sustained use of the item bank. The item bank cannot be successful and sustainable unless it is supported and valued by educators.

REFERENCES CITED

Gunzenhauser, M. (2003). High-Stakes Testing and the Default Philosophy of Education. *Theory into Practice,* 42.

Guskey, T. R. (2007). Multiple sources of evidence: An analysis of stake-holders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice,*19–27.

Hsu, T. (2005). Research Methods and Data Analysis Procedures Used by Educational Researchers. *International Journal of Research & Method in Education, 28.*

Jones, B. D. & Egley, R. J. (2004). Voices from the Frontlines: Teachers' Perceptions of High-Stakes Testing. *Education Policy Analysis Archives,* 12 (39).

Osburn, M., Stegman, C., Suitt, L., & Ritter, G. (2004). Parents' Perceptions of Standardized Assessment: Its Relationship and Effect on Student Achievement. *Journal of Educational Research and Policy Studies,* 4, 75–95.

Reeves, D. B. (2007). From the Bell Curve to the Mountain: A New Vision for Achievement, Assessment, and Equity. In D.B. Reeves (Ed.), *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning* (pp. 1–11 and 227–249). Bloomington: Solution Tree Press

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment *for* learning: Impact on student achievement. *Assessment in Education: Principles, Policy, and Practice,* 11(1), 49–65

APPENDICES

APPENDIX A

ITEM WRITER SURVEY

Item Writer Survey
Florida Interim Assessment Item Bank and Test Platform

September 2012
Please indicate the extent to which you agree with each of the following statements.

___ Strongly Disagree
___ Somewhat Disagree
___ Somewhat Agree
___ Strongly Agree

1. In general, the information presented in this training increased your understanding of item writing.
2. Training increased your understanding of the parts of an item.
3. Training included clear definitions of the basic terminology used in item development.
4. Training helped you understand the importance of strong distractors.
5. Training provided enough information on how to determine the cognitive difficulty of a test item.
6. Training aided your understanding of how the principles of accessibility and Universal Design can be applied in item writing.
7. Training increased your understanding of how to create an item stimulus.
8. Training increased your understanding of how to create selected response items.
9. Training increased your understanding of how to create open-ended response items.
10. Training increased your understanding of how to create rubrics.
11. The training event was well organized.
12. The sequence of training topics was logical and easy to follow.
13. The training facilitator seemed to be well prepared.
14. As a result of participating in training, you feel better prepared to write high-quality test items.
15. Is there anything you would like to tell us about your experience in this training event?

APPENDIX B

ITEM REVIEWER SURVEY

Item Reviewer Survey
Florida Interim Assessment Item Bank and Test Platform

September 2012
Please indicate the extent to which you agree with each of the following statements.

___ Strongly Disagree
___ Somewhat Disagree
___ Somewhat Agree
___ Strongly Agree

1. In general, the information presented in this training increased your understanding of item reviewing.
2. Training increased your understanding of reviewing the parts of an item.
3. Training included clear definitions of the basic terminology used in item development and item review.
4. Training helped you understand the importance of strong distractors.
5. Training provided enough information on how to determine the cognitive difficulty of a test item in an effective review.
6. Training aided your understanding of how the principles of accessibility and Universal Design can be applied in item writing and item review.
7. Training increased your understanding of how to review an item stimulus.
8. Training increased your understanding of how to review selected response items.
9. Training increased your understanding of how to review open-ended response items.
10. Training increased your understanding of how to review rubrics.
11. The training event was well organized.
12. The sequence of training topics was logical and easy to follow.
13. The training facilitator seemed to be well prepared.
14. As a result of participating in training, you feel better prepared to review test items.
15. Is there anything you would like to tell us about your experience in this training event?

APPENDIX C

SCIENCE EDUCATOR SURVEY

Science Educator Survey
Florida Interim Assessment Item Bank and Test Platform

December 2012

This survey is intended to collect information about educators' knowledge and perceptions of the Florida Interim Assessment Item Bank and Test Platform. The information will be used by the Office of Race to the Top Assessments to improve the quality and utility of the Item Bank. Participation in this survey is voluntary. Please contact Sally Sanders, sally.sanders@fldoe.org, if you have questions or comments.

1. What is the size of your school district? Use the chart which indicates school district size based on student population.
   Small___ Medium___ Large___

2. How many years of science education experience do you have (including district-level)?
   <6____ 6-10____ 11-15____ 16-20____ >20____

3. Does your district have science teachers serving as science item writers or item reviewers for the Item Bank?
   Writers and Reviewers___ Writers only___ Reviewers only___
   Neither___

   For questions 4 – 8, please indicate the extent to which you agree with each statement.

4. My district is planning to provide professional development related to the Item Bank.
   3)Strongly agree___ 2)Somewhat agree___ 1) Disagree___

   Comments:

5. My district will use the Item Bank for science assessment. (Please note that use of the Item Bank is independent of FCAT and EOC testing requirements).
   3) Strongly agree ___ 2)Somewhat agree___ 1) Disagree___

Comments:

6. The Item Bank will be a valuable and useful tool for meeting my district's needs for standards-based testing.
   3) Strongly agree___ 2)Somewhat agree___ 1)Disagree___

   Comments:

7. Other science teachers in my district view the Item Bank as a valuable and useful tool for science assessment.
   3) Strongly agree___ 2) Somewhat agree___ 1) Disagree___

   Comments:

8. My district is interested in contributing items to the Item Bank.

   3) Strongly agree___ 2) Somewhat agree___ 1) Disagree____

   Comments:

9. What are your district's needs for assistance from the Office of Race to the Top Assessments related to the Item Bank?

10. Other comments, questions, or suggestions:

Thank you for participating in the survey!

APPENDIX D

MONTANA STATE UNIVERSITY
INSTITUTIONAL REVIEW BOARD
EXEMPTION LETTER

MONTANA STATE UNIVERSITY

INSTITUTIONAL REVIEW BOARD
For the Protection of Human Subjects
FWA 00000165

960 Technology Blvd. Room 127
c/o Immunology & Infectious Diseases
Montana State University
Bozeman, MT 59718
Telephone: 406-994-6783
FAX: 406-994-4303
E-mail: cherylj@montana.edu

Chair: Mark Quinn
406-994-5721
mquinn@montana.edu
Administrator:
Cheryl Johnson
406-994-6783
cherylj@montana.edu

MEMORANDUM

TO: Sally Sanders and Walt Woolbaugh

FROM: Mark Quinn, Chair Mark Quinn CJ

DATE: December 7, 2012

RE: *How Can the Functionality and Value of the Florida Item Bank and Test Platform be Maximized? [SS120712.A-EX]*

The above research, described in your submission of December 7, 2012, is exempt from the requirement of review by the Institutional Review Board in accordance with the Code of Federal regulations, Part 46, section 101. The specific paragraph which applies to your research is:

X (b) (1) Research conducted in established or commonly accepted educational settings, involving normal educational practices such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

X (b) (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

___ (b) (3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if: (i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.

___ (b) (4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available, or if the information is recorded by the investigator in such a manner that the subjects cannot be identified, directly or through identifiers linked to the subjects.

___ (b) (5) Research and demonstration projects, which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (i) public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.

___ (b) (6) Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed, or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the FDA, or approved by the EPA, or the Food Safety and Inspection Service of the USDA.

Although review by the Institutional Review Board is not required for the above research, the Committee will be glad to review it. If you wish a review and committee approval, please submit 3 copies of the usual application form and it will be processed by expedited review.

APPENDIX E

ITEM WRITER SURVEY RESULTS

Florida IBTP Training Survey Results

1. In general, the information presented in this training increased your understanding of item writing.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	0	0%
Somewhat Agree	17	44%
Strongly Agree	21	54%
Total	39	100%
Average	3.4871795	

2. Training increased your understanding of the parts of an item.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	0	0%
Somewhat Agree	12	31%
Strongly Agree	26	67%
Total	39	100%
Average	3.6153846	

3. Training included clear definitions of the basic terminology used in item development.	# of Responses	% of Responses
Strongly Disagree	0	0%
Somewhat Disagree	1	3%
Somewhat Agree	11	28%
Strongly Agree	27	69%
Total	39	100%
Average	3.6666667	

4. Training helped you understand the importance of strong distractors.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	0	0%
Somewhat Agree	10	26%
Strongly Agree	28	72%
Total	39	100%
Average	3.6666667	

5. Training provided enough information on how to determine the cognitive difficulty of a test item.	# of Responses	% of Responses
Strongly Disagree	0	0%
Somewhat Disagree	4	10%
Somewhat Agree	23	59%
Strongly Agree	12	31%
Total	39	100%
Average	3.2051282	

6. Training aided your understanding of how the principles of accessibility and Universal Design can be applied in item writing.	# of Responses	% of Responses
Strongly Disagree	0	0%
Somewhat Disagree	4	10%
Somewhat Agree	13	33%
Strongly Agree	22	56%
Total	39	100%
Average	3.4615385	

7. Training increased your understanding of how to create an item stimulus.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	2	5%
Somewhat Agree	16	41%
Strongly Agree	20	51%
Total	39	100%
Average	3.4102564	

8. Training increased your understanding of how to create selected response items.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	1	3%
Somewhat Agree	15	38%
Strongly Agree	22	56%
Total	39	100%
Average	3.4871795	

9. Training increased your understanding of how to create open-ended response items.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	3	8%
Somewhat Agree	18	46%
Strongly Agree	17	44%
Total	39	100%
Average	3.3076923	

10. Training increased your understanding of how to create rubrics.	# of Responses	% of Responses
Strongly Disagree	0	0%
Somewhat Disagree	6	15%
Somewhat Agree	24	62%
Strongly Agree	9	23%
Total	39	100%
Average	3.0769231	

11. The training event was well organized.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	0	0%
Somewhat Agree	10	26%
Strongly Agree	28	72%
Total	39	100%
Average	3.6666667	

12. The sequence of training topics was logical and easy to follow.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	0	0%
Somewhat Agree	11	28%
Strongly Agree	27	69%
Total	39	100%
Average	3.6410256	

13. The training facilitator seemed to be well-prepared.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	1	3%
Somewhat Agree	7	18%
Strongly Agree	30	77%
Total	39	100%
Average	3.6923077	

14. As a result of participating in training, you feel better prepared to write high-quality test items.	# of Responses	% of Responses
Strongly Disagree	1	3%
Somewhat Disagree	1	3%
Somewhat Agree	17	44%
Strongly Agree	20	51%
Total	39	100%
Average	3.4358974	

15. Is there anything you would like to tell us about your experience in this training event?

APPENDIX F

ITEM REVIEWER SURVEY RESULTS

2012 Florida IBTP Item Reviewer Survey Results

1. In general, the information presented in this training increased your understanding of item reviewing.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	1	1%
Somewhat Agree	26	36%
Strongly Agree	44	61%
Total	72	100%
Average	3.5694444	

2. Training increased your understanding of reviewing the parts of an item.	# of Responses	% of Responses
Strongly Disagree	2	3%
Somewhat Disagree	0	0%
Somewhat Agree	26	36%
Strongly Agree	44	61%
Total	72	100%
Average	3.5555556	

3. Training included clear definitions of the basic terminology used in item development and item review.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	0	0%
Somewhat Agree	9	12%
Strongly Agree	62	86%
Total	72	100%
Average	3.8333333	

4. Training helped you understand the importance of strong distractors.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	0	0%
Somewhat Agree	23	32%
Strongly Agree	48	67%
Total	72	100%
Average	3.6388889	

5. Training provided enough information on how to determine the cognitive difficulty of a test item in an effective review.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	7	10%
Somewhat Agree	37	51%
Strongly Agree	27	38%
Total	72	100%
Average	3.25	

6. Training aided your understanding of how the principles of accessibility and Universal Design can be applied in item writing and item review.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	4	6%
Somewhat Agree	31	43%
Strongly Agree	36	50%
Total	72	100%
Average	3.4166667	

7. Training increased your understanding of how to review an item stimulus.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	4	6%
Somewhat Agree	26	36%
Strongly Agree	41	57%
Total	72	100%
Average	3.4861111	

8. Training increased your understanding of how to review selected response items.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	0	0%
Somewhat Agree	23	32%
Strongly Agree	48	67%
Total	72	100%
Average	3.6388889	

9. Training increased your understanding of how to review open-ended response items.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	1	1%
Somewhat Agree	26	36%
Strongly Agree	44	61%
Total	72	100%
Average	3.5694444	

10. Training increased your understanding of how to review rubrics.	# of Responses	% of Responses
Strongly Disagree	2	3%
Somewhat Disagree	2	3%
Somewhat Agree	32	44%
Strongly Agree	36	50%
Total	72	100%
Average	3.4166667	

11. The training event was well organized.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	0	0%
Somewhat Agree	28	39%
Strongly Agree	43	60%
Total	72	100%
Average	3.5694444	

12. The sequence of training topics was logical and easy to follow.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	3	4%
Somewhat Agree	12	17%
Strongly Agree	56	78%
Total	72	100%
Average	3.7083333	

13. The training facilitator seemed to be well-prepared.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	0	0%
Somewhat Agree	15	21%
Strongly Agree	56	78%
Total	72	100%
Average	3.75	

14. As a result of participating in training, you feel better prepared to review test items.	# of Responses	% of Responses
Strongly Disagree	1	1%
Somewhat Disagree	0	0%
Somewhat Agree	26	36%
Strongly Agree	45	62%
Total	72	100%
Average	3.5972222	