



## Request for Offers

### Independent Verification of the Psychometric Validity for the Florida Standards Assessment

### Technical and Cost Proposal

**Submitted to:**

Attn: Ms. Ashley Spicola  
1602 The Capitol  
Tallahassee, FL 32399-0400

**Point of Contact, Submitted by:**

Chad W. Buckendahl, Ph.D.  
Director of Strategic Partnerships  
Alpine Testing Solutions  
51 W. Center Street, #514  
Orem, UT 84057  
Phone/Fax: (702) 586-7386  
Mobile: (402) 770-0085  
Email: [chad.buckendahl@alpinetesting.com](mailto:chad.buckendahl@alpinetesting.com)  
[www.alpinetesting.com](http://www.alpinetesting.com)

E-Signature: *Chad Buckendahl*

Date: May 26, 2015

# Table of Contents

|  |    |
|--|----|
| Overview .....                                     | 3  |
| Corporate Capacity .....                           | 4  |
| Alpine Testing Solutions.....                      | 4  |
| edCount, LLC .....                                 | 5  |
| Technical Plan.....                                | 6  |
| a. Project Plan.....                               | 6  |
| b. Evaluations.....                                | 7  |
| Evaluation Design.....                             | 7  |
| Evaluation of Test Items .....                     | 11 |
| Evaluation of Field Testing.....                   | 11 |
| Evaluation of Test Blueprint and Construction..... | 12 |
| Evaluation of Test Administration.....             | 12 |
| Evaluation of Scaling, Equating, and Scoring ..... | 13 |
| Specific Evaluation of Psychometric Validity ..... | 14 |
| c. Project Reporting .....                         | 15 |
| Preliminary Evaluation Reports .....               | 15 |
| Final Report .....                                 | 15 |
| Proposed Project Schedule.....                     | 16 |
| Management and Staffing Plan .....                 | 17 |
| Organizational Chart .....                         | 18 |
| Alpine Testing Solutions.....                      | 19 |
| edCount, LLC .....                                 | 20 |
| Summary .....                                      | 21 |
| References.....                                    | 22 |
| Appendix A: Resumes of Key Staff.....              | 23 |
| Attachment 1 – Price Sheet.....                    | 40 |
| Attachment 2 – Disclosure Statement.....           | 42 |

## Overview

Alpine Testing Solutions (Alpine), in partnership with edCount, LLC (edCount) are pleased to offer this proposal to the Florida Department of Education (FLDOE) to conduct an independent verification of the psychometric validity for the Florida Standards Assessment program. The partners submitting this proposal are dedicated to the improvement of educational assessment by providing research and consultation services to state departments of education, testing agencies, and other organizations to assist in their test development and validation efforts.

Staff members at Alpine have conducted independent evaluations and psychometric audits for more than 15 years, including leading the most recent Evaluation of the *National Assessment of Educational Progress* (Buckendahl, et al., 2009), conducting an external evaluation of secondary assessment for the Wyoming State Board of Education, conducting an external psychometric and security audit for Western Governors University, providing external psychometric evaluation and quality control services for Oklahoma's educator licensure program, conducting a psychometric audit of the National Commission for the Certification of Physician Assistants, and conducting semi-annual evaluations of the Florida Bar Exam for the Florida Board of Bar Examiners.

Similarly, in consultation with state departments of education and multi-state consortia, edCount has provided assessment design, policy consultation, and also conducted independent evaluations of assessment programs that result in actionable recommendations, including the National Center and State Collaborative funded by the Office of Special Education Programs at the US Department of Education and involving a consortia of 24 states and entities. For this on-going project, edCount focuses specifically on the validity evaluation of the developing alternate assessment as well as related curricula and professional development modules.

Further, edCount has conducted numerous studies of item alignment quality and other validity and technical documentation issues with consortia and individual states across the nation (e.g., Alaska, Arizona, Connecticut, Delaware, Florida, Georgia, Hawaii, Indiana, Louisiana, Massachusetts, Mississippi, Montana, Nebraska, Nevada, New York, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Washington, Wisconsin, Wyoming, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and the Pacific Assessment Consortium (PAC-6), PARCC). edCount applied these same skills in the federally-funded project, Evaluating the Validity of English Language Proficiency Assessments, involving Washington state, Oregon, Idaho, Montana, and Indiana. Collectively, the evaluation team represents some of the most experienced professionals in educational research, validation, and the evaluation of testing programs.

A valuable characteristic of the organizations in this partnership is their notable independence from some of the real and perceived conflicts of interest that can jeopardize such evaluations. Specifically, neither Alpine, nor edCount develop PK-12 educational assessments that target the grade levels, content areas, or courses within the scope of this evaluation project. In addition, neither Alpine, nor edCount have served in a prior capacity to advise or evaluate Florida assessments, so we would not be in a position of potentially evaluating the outcomes of our own

recommendations for practice. As a result, we believe that we can offer high quality evaluation services that will yield credible, actionable recommendations for the Florida Standards Assessments (FSA) program.

In addition to our independence and expertise in assessment design, implementation, and related policy, there are two other critical elements of our partnership that are important to note. First, Alpine and edCount envision consistent and close contact with FLDOE staff throughout the design, implementation, and reporting activities of the evaluation. The information that is gathered during the project and the evaluation process for all steps will be transparent to the FLDOE and its constituents. Our proposed design represents the methods and procedures we believe to be an appropriate and valid way to independently evaluate the FSA within the expected scope defined in the solicitation. We believe our design will allow our organizations to provide the FLDOE a comprehensive review of its assessment characteristics along with the overall assessment program, and will also support the delivery of actionable information that will present not only a report that highlights the strengths and weaknesses of their program, but also concrete steps that the FLDOE can follow to improve the health of the assessment program.

Second, our proposed solution emphasizes collaboration among the complementary expertise of Alpine and edCount with each organization taking leadership roles within the project to focus on their respective strengths. Throughout the development and implementation of the evaluation design, leadership from each organization will work collaboratively to conduct the evaluation studies, discuss the results, create reports, and produce meaningful and actionable deliverables.

## Corporate Capacity

### Alpine Testing Solutions

Alpine Testing Solutions (Alpine) is an employee-owned company that focuses on psychometric consultation, test development, data analysis, and credential management. Alpine works with clients across the education (PK-12, admissions, adult education, language testing), and credentialing sectors (licensure, certification, registration, and assessment-based certificates).

Alpine staff members' collective understanding of the interrelated elements of test development, policy, and psychometrics, coupled with their ability to apply those principles in the context of each specific testing program, sets us apart from other organizations in the testing community. Key staff members have provided a range of services—including general psychometric consultation within the context of applicable policy; consulting on program design; evaluating and auditing programs; developing validation frameworks; and conducting test development and validation projects (e.g., practice/job analysis, blueprint development, item development, item and form analysis, forms assembly, standard setting and equating, security analysis)—to small- and large-scale testing programs of national and international reach that use a variety of administration models. Many of these programs rely on the use of written tasks and performance tasks to best represent the content, cognitive demand, performance demand, and environment necessary for the domain.

As an organization, Alpine has worked with a variety of clients in the educational market, as well as in the professional certification and licensure field. As noted previously, Alpine's staff members have

designed and implemented a number of psychometric evaluations of educational and credentialing testing programs for national (e.g., *National Assessment of Educational Progress*, Western Governors University, National Commission for Certification of Physician Assistants) and state level programs (e.g., Wyoming State Board of Education, Florida Board of Bar Examiners, Oklahoma Office of Educational Quality and Accountability). In addition, senior staff members proposed to lead this project have contributed to the professional community specifically on the topic of assessment program evaluation (e.g., Wiley, 2015; Buckendahl, 2015; Buckendahl, Plake, & Davis, 2009; Buckendahl & Plake, 2006). As a result, the Alpine team is uniquely qualified to collaborate with edCount and FLDOE on this project.

### edCount, LLC

edCount, LLC, is a federally-registered woman-owned small business and a certified Woman-Owned Business Enterprise. Since its founding in 2003, edCount has provided direct or advisory services in K-12 assessment to all 50 states and seven U.S. territories through projects funded via both competitive and sole source opportunities ranging from \$10,000 to over \$3,000,000 annually. As an independent small business that does not offer operational testing services, edCount is able to offer state education agencies objective, constructive feedback about their assessment programs.

edCount staff members have extensive experience assisting state and local education agencies with the evaluation of their assessment systems. edCount has designed and implemented dozens of studies involving content, internal structure, external relationship, response process, and consequential sources of validity evidence and provide clients with clear and concise reports in every case. edCount staff are adept in both quantitative and qualitative forms of assessment evaluation and at combining evidence to address specific evaluation questions. Further, its experience with state and local education agency staff allow us unique insights into the demands its clients face and their needs to communicate complex information to a wide-range of stakeholders. As a result, edCount ensures that reports and documentation are accurate, accessible, and designed to address the specific purposes the clients define.

In addition to the many opportunities to serve US states and school districts, edCount has had as clients or partners a wide range of entities that recognize edCount's contributions to improving assessment quality across the country. These entities include the U.S. Department of Education (including the offices of English Language Acquisition, Planning, Evaluation and Policy Development, Special Education and Rehabilitative Services, Elementary and Secondary Education, and the National Center for Education Statistics); The Laurent Clerc Center at Gallaudet University; The National Alternate Assessment Center at the University of Kentucky; The Education Alliance at Brown University; Keystone Alternate Assessment Design; Pearson; Thompson Publishing; the Council of Chief State School Officers; the National Center for the Improvement of Educational Assessments; the National Center on Educational Outcomes at the University of Minnesota; the Graduate School of Education and Information Studies at the University of California, Los Angeles; the College of Education at the University of North Carolina, Charlotte; the Pacific Institute for Research and Evaluation; ICF International; Westat; Education Northwest; and Mathematica Policy Research.

## Technical Plan

All services proposed herein by Alpine Testing Solutions (Alpine) and edCount, LLC (edCount) are grounded in a validity-centered approach to test development, scoring, and reporting. *The Standards for Educational and Psychological Testing* (2014), jointly published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), are the primary guidelines used to evaluate current practices of testing programs to ensure they are consistent with the standards of the professional testing community. In addition, we will rely on supplementary resources such as *Educational Measurement*, 4<sup>th</sup> ed. (Brennan, 2006); and the *Handbook of Test Development* (Downing & Haladyna, 2006) to help ensure programs are consistent with industry best practices.



Figure 1: Test Development and Validation Process

Figure 1 illustrates the general approach to test development and validation recommended by the measurement community. Validity – the degree to which theory and evidence supports the intended uses and interpretations of test scores – is the central focus. By approaching test development and maintenance in this way, each task and its related activities provides an opportunity to evaluate validity and document evidence in support of the program. Note that administration to collect information can be subsumed within the pre-test and analyze and/or at the operational phase once test forms are created. Our team’s validity-centered approach to evaluating test development and maintenance is well-aligned to the goals and evaluation activities sought by FLDOE.

### a. Project Plan

At the outset of the project, the evaluation team will schedule a kickoff meeting to discuss the proposed approach to ensure that there is a mutual understanding of expectations and outcomes. These will be formalized as a Project Plan that describes the specific activities for each evaluation inquiry along with, the sources of evidence requested, and a timeline for completing these activities that aligns with the larger schedule for the project.



The project plan will be delivered to the FLDOE within fifteen (15) working days of the contract award and include a clear set of measurable milestones for each project that will track the progress of each progress to completion. Throughout the life of the contract, weekly conference calls will be scheduled with the FLDOE. Prior to each weekly conference call, an agenda will be shared with the FLDOE, and minutes from each call will be shared with the FLDOE within three (3) working days. During these calls, progress will be reported to the FLDOE using these milestones to allow all parties to fully comprehend how each milestone is moving forward.

As part of this project plan, Alpine will collaborate with FLDOE to develop a structure for the project that will provide timely and accurate information while adhering to requirements described in the solicitation. A system for communication between the FLDOE, its vendor, and the evaluation team will be designed to ensure that FLDOE staff are able to easily gather information on the status and upcoming work for all of the tasks associated with this project. To facilitate the communication we are proposing to support a technology infrastructure to maintain an open conduit. Specifically, tracking documents will be maintained in real time, and will be made available to FLDOE staff in an agreed upon location (e.g., FTP site, secure FLDOE-specific Sharepoint site) so the status of the project can be accessed and reviewed at any time.

In addition to providing access to status reports in real time, Alpine will develop a system for the simple communication with the FLDOE if any unexpected obstacles arise for any portion of the project. In the event that such obstacles were to arise, the FLDOE will be promptly informed about the nature of the obstacle, the implications for the overall schedule, and the options for overcoming the obstacle.

## b. Evaluations

In this section we propose a comprehensive evaluation plan based on the targeted evaluations requested by FLDOE. Our work will begin by analyzing the intended interpretations and uses of scores in addition to the theory of action associated with the FSA to identify the critical features that should be considered in the project. When evaluating evidence for the elements of the Florida Standards Assessment program, our team will rely on the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), using the five defined sources of validity evidence as an expected validation framework. Specifically, validity evidence based on test content, response processes, internal structure, relations to other variables, and testing consequences.

## Evaluation Design

Our proposed evaluation design is based predominantly on evidence that has been developed or collected throughout the test development, validation, delivery, and analyses processes for the FSA. The general process for reviewing the various components will involve a professional review of the individuals or groups involved, the processes and procedures that occurred, the results of those processes, decision rules that were applied, and decisions that emerged. These elements of the review are consistent with professional expectations of the *Standards* and also with best practice (see, for example, Buckendahl & Plake, 2006).

A proposed validation framework is presented in Table 1. Descriptions of each of these studies are then described. Note that not all cells in Table 1 are populated. This can be explained by how the sources evidence are used in practice. The evaluation design only draws from those areas most relevant for a comprehensive validation argument. It does not require evidence from every source for every target area. Also note that some evidence may not currently be available or cannot be collected within the timeframe allocated for this project.



Table 1. Proposed Validation Framework for Independent Verification of Psychometric Validity of Florida Standards Assessment.

| Evaluation Target Areas                       | AERA et al. (2014) Source of Validity Evidence   |   |  |  |   |
|---|--|---|--|--|---|
|   | Test Content   | Response Processes  | Internal Structure                                   | Relations to other Variables   | Testing Consequences  |
| Evaluation of Test Items                      | Review test development and review processes<br><br>Review sample of assessment items for content and potential bias | Review student and grade level language; cognitive levels |  |  | Review based on research-based instructional methodology  |
| Evaluation of Field Testing                   |  |   | Review rationale, execution, and results of sampling |  | Review whether results support test construction  |
| Evaluation of Test Blueprint and Construction | Review test blueprint for sufficiency to support intended purposes   |   |  |  | Review the utility of score reports for stakeholders to improve instruction   |
| Evaluation of Test Administration             |  | Review of test accommodations                             |  | Review of delivery system utility and user experience<br><br>Review of third party technology and security audit reports | Review of test administration procedures<br><br>Review of security protocols for prevention, investigation, and enforcement |

|   |   |  |  |  |  |
|---|---|--|--|--|--|
| <p>Evaluation of Scoring, Scaling, and Equating</p> | <p>Review evidence of content validity produced by the program</p>                        | <p>Review evidence of content validity produced by the program</p>                       | <p>Review choice of model, scoring, analyses, equating, and scaling.</p> <p>Subgroup psychometric characteristics</p> <p>Subscore added value analyses, decision consistency and measurement precision</p> | <p>Review evidence of construct validity collected by the program</p> <p>Review criterion evidence collected by the program.</p> | <p>Review evidence of testing consequences produced by the program</p> |
| <p>Specific Evaluation of Psychometric Validity</p> | <p>Review a sample of items relative to course descriptions and for freedom from bias</p> | <p>Review of a sample of items for intended response behavior as opposed to guessing</p> | <p>Review of item difficulty, discrimination, potential bias</p> <p>Review the linking processes for Algebra 1 and Grade 10 ELA relative to 2013-14 results.</p>   |  |  |

## Evaluation of Test Items

To evaluate evidence of content, response processes, and testing consequences, the design and implementation of this evaluation study will focus on how the assessment development and review processes align with professional expectations. Given their experience in conducting alignment studies and designing assessments, staff members from edCount will lead this evaluation study. Specifically, they will review documentation regarding a series of development activities using criteria based on best practices in the industry. In addition, staff members will conduct in-person and virtual interviews with FLDOE and its vendor, AIR, to gather information not included in documentation or to clarify evidence. The following elements of the study with illustrative review criteria will be included:

- Review test development and review processes
  - Qualified subject matter experts familiar with content and target students
  - Systematic review processes with appropriate quality control were implemented
  - Results were consistent with expected outcomes of the processes with any changes that were recommended during the review processes
  - Decision rules about item development and review that yielded the results
- Review a minimum of 200 operational assessment items across grades and content areas
  - Structured consistently with best practices in assessment item design
  - Consistent with widely accepted, research based instructional methods
  - Appropriate cognitive levels to target intended depth of knowledge
  - Review for potential bias related to sex, race, ethnicity, socioeconomic status
  - Appropriate student and grade level language
  - Targeting the intended content standard(s)

## Evaluation of Field Testing

The evaluation of the field testing procedures is a critical component of this project because the data collected during this process will provide evidence that impacts all components of the assessment program. Specifically, if the data collected during the field test is suspect, it can threaten the validity of the psychometric analyses, establishing of cut scores, equating, and scaling. Alpine will lead this study and focus on the test development and validation processes followed during the field testing. To accomplish this, the evaluation will focus on two aspects of the validation framework; specifically, internal response processes and testing consequences. Staff members will review documentation and data from these activities, supplementing this information with interviews with FLDOE and AIR staff. The elements and review criteria that the team will apply are noted here:

- Review the sampling plan
  - Review the design for consistency with intended purpose(s)
  - Review the process for creating the sampling plan
  - Review the extent to which the sampling plan was implemented
  - Review any processes and procedures to ensure evidence of sufficient sample size and population representation
  -
- Review whether results of the sampling support test construction
  - Review the sufficiency of information and stability of field test results to support test form construction

- Review whether the field test results yield results support a range of raw scores that would be transformed into scale scores relative to cut scores
- Review and decision rules that were applied to the results of the field test
- Review the capacity of the assessments to support intended score reports and related assessment data to inform instruction (e.g., group level, individual student level)

### Evaluation of Test Blueprint and Construction

To evaluate evidence of content and testing consequences, the design of this evaluation study will focus on how the test blueprint and construction process are consistent with the intended interpretations and uses of scores from the assessment. In addition, this study will also evaluate whether the blueprint has the capacity to support the intended reports for the program. Staff members from edCount will lead this evaluation study, reviewing documentation regarding the blueprint development and test construction activities using criteria based on industry best practices. In addition, staff members will conduct in-person and virtual interviews with FLDOE and AIR to gather information not included in documentation or to clarify evidence. The following elements of the study with illustrative review criteria will be included:

- Review test blueprint for sufficiency of intended purposes
  - Review the design of the blueprint relative to intended interpretations and uses
  - Review sufficiency of information for domain or subscores intended for reporting to stakeholder groups
- Review the utility of score reports for stakeholders
  - Review of design of score reports for stakeholder groups
  - Review of graphics – static, dynamic – associated with score reports
  - Review of explanatory text for appropriateness to intended population
  - Review of score reports for information to support improvement of instruction

### Evaluation of Test Administration

Given some of the challenges that were publicly reported regarding administration of the FSA, an evaluation of the test administration practices will contribute important information about not only the design and implementation of the delivery platform, but also the potential impact on the validity of scores for students in Florida. Staff members from Alpine will lead this study to focus on the elements of response processes, relations with other variables, and testing consequences within the validation framework recommended by the industry standards. To conduct this evaluation study, staff members will review five sources of evidence through a review of documentation, reviewing the delivery system, and conducting in-person and virtual interviews with staff at FLDOE, AIR, and selected local education agencies who administered the assessments. The organization of this study and expected review criteria are noted here:

- Review of test accommodations
  - Review of type and utilization of accommodations during administration
  - Review of universal design principles applied to the delivery system
- Review of delivery system utility and user experience from local education agencies
  - Review of the implementation of the delivery system with recommended technical specifications

- Review of research conducted while creating the user interface during the development of the delivery system
- Review of local education agencies experience with the delivery system and user experience
- Review of third party technology and security audit reports
  - Review of relevant external audits of the delivery system (e.g., stress testing, penetration testing)
- Review of test administration procedures
  - Review of system requirements for delivery system as applied to local education agencies
  - Review of documented procedures for administering the assessments
- Review of security protocols for prevention, investigation, and enforcement
  - Review of documented procedures for security protocols during assessment administration
  - Review of reports regarding detection and investigation of security concerns
  - Review of reports of anomalies during assessment administration
  - Reports of enforcement policies and practices for assessment administration

### Evaluation of Scaling, Equating, and Scoring

This evaluation study is intended to provide information about whether the scoring, equating, and scaling activities were completed within expected industry standards. Because this element of the scope includes a request to review content, criterion, construct, and consequential evidence, this study will target all five aspects of the validation framework. In some instances, it is possible that evidence from FLDOE or AIR may not yet be available during the targeted timeframe of the study. If so, we will review the plans for collecting and evaluating this evidence in the future. Staff members from Alpine will lead the review of these areas, focusing on each of the elements in the validation framework. In conducting this study, staff members will review seven sources of evidence through a review of documentation and conducting in-person and virtual interviews with staff at FLDOE and AIR. The organization of this study and review criteria are noted here:

- Review evidence of content validity collected by the program
  - Review evidence of qualified subject matter experts
  - Review evidence of appropriate processes and procedures
  - Review evidence of results that support claims of content validity
- Review rationale for scoring model, analyses, equating, and scaling
  - Review evidence that supports the choice of the scoring model
  - Review the implementation and results of the psychometric analyses
  - Review the design, implementation, results, and decision rules for equating
  - Review the design, implementation, results, and decision rules for scaling for total scores and domain or subscores
- Review psychometric characteristics of the assessments
  - Review analyses of reliability, inclusive of standard error of measurement
  - Review decision consistency and accuracy
  - Review subscore added value analyses
- Review psychometric characteristics of subgroups

- Review psychometric performance of assessment items for reporting subgroup performance (e.g., reliability of subgroups, differential item functioning)
- Review evidence of construct validity collected by the program
- Review evidence of criterion validity collected by the program
  - Review identified criterion variables and related studies
- Review evidence of testing consequences collected by the program
  - Review evidence that may include impact on curriculum and instruction, impact on student achievement, impact on teacher behavior, and public perception

### Specific Evaluation of Psychometric Validity

To evaluate the specific elements of psychometric validity identified in this section of the scope, staff members from edCount and Alpine will co-lead this study to take advantage of their complementary skills. The evaluation questions in this section target content, response processes, and internal structural components of the validation framework. The collective team will review documentation regarding development activities using criteria based on best practices in the industry. To supplement the information contained in documentation, staff members will conduct in-person and virtual interviews with FLDOE and AIR to gather information not included in documentation or to clarify evidence. The following elements of the study with illustrative review criteria will be included:

- Review a sample of items from each grade and subject
  - Review items for content, cognitive processes, and performance levels relative to standards as described in course descriptions
  - Review items for design characteristics that reduce the likelihood that the student answers the question correctly by guessing
  - Review evidence of the fairness or bias review
- Review psychometric characteristics of items
  - Review item difficulty results for an acceptable range of parameters
  - Review item discrimination results for an acceptable range of parameters
  - Review option analyses for functional item response characteristics
  - Review empirical evidence of potential bias such as differential item functioning
- Review the linking processes for Algebra 1 and Grade 10 ELA to 2013-14 results
  - Review the assumptions for the linking studies
  - Review the design of the linking studies
  - Review the results and associated decision rules applied in the linking studies
  - Review the communication reports regarding the linking and the information to schools and other Florida constituents

An important feature of our proposed design is the understanding that not all assessment programs are developed following the exact same protocols and procedures. Instead, each program has unique requirements, and as a result, our team's approach to the FSA independent verification will be program-specific. Based on our extensive experience with various operational programs across industries, we understand that the creation of a valid assessment program can apply different, yet appropriate methods that yield quality assessments and meet industry standards. Through document review, interviews with program, vendor, and field staff, we will gain the understanding of the program needed to provide an accurate and detailed review to be able to provide specific commendations and recommendations improvement.

### c. Project Reporting

The project team is fully committed to the reporting expectations outlined in the solicitation – status reports, preliminary evaluation reports, and the final report. Status reports will be produced monthly to provide an update on the project’s progress. The preliminary and final evaluation reports are described in more detail in the next sections. Reports that are drafted will undergo both a technical review by the project’s management team and a style review by Alpine’s Managing Editor to ensure that the reports are accessible to a range of stakeholders. FLDOE will have an opportunity to review draft reports produced for this project for factual accuracy; after which the project team will make any necessary revisions and deliver the Final Report as the final deliverable.

#### Preliminary Evaluation Reports

Reports will be provided at the preliminary stage for each of the evaluation studies described. Each summary will include the following components of the respective evaluation:

- Executive summary of the evaluation study (e.g., Test Development)
- Background data and information to explain the value and importance
- The specific questions being investigated
- A review of the relevant literature that informed the study
- The framework for the study design
- Data collection procedures
- Analysis (either planned or completed)
- Results (if available)
- Interpretive or contextual information to help interpret the results
- Limitations of the results or the study design
- Preliminary conclusions, commendations, or recommendations from the study

#### Final Report

The final evaluation report will represent a comprehensive review of all evaluation studies completed as part of the project, along with the resulting conclusions, commendations, and recommendations. The comprehensive final report will include a synthesized executive summary and description of the activities that occurred, the studies that were included, results, and recommendations. In addition, for every study completed, the study report will include:

- Executive summary of the study
- Background data and information to explain the value and importance
- The specific questions that were investigated
- A review of the relevant literature that informed the study
- The framework for the overall study design
- Data collection procedures
- Analyses that were completed
- Results



- Interpretive or contextual information to help interpret the results
- Limitations of the results or the study design
- Conclusions, commendations, or recommendations from the study

The final report will also include a review of any outstanding issues or concerns that were identified during the project that could not be addressed within the timeframe of the project, along with recommendations for how they could be investigated.

Through the combination of studies proposed herein, Alpine and edCount are confident that we can provide the independent, high quality independent verification of the psychometric validity of the FSA requested by the FLDOE.

## Proposed Project Schedule

The following schedule of activities is proposed based on our interpretation of the scope of work and will be negotiated with FLDOE during the project planning phase. Deliverables are noted with an asterisk (\*).

| <u>Activity</u>                                  | <u>Dates</u>       | <u>Responsibility</u>  |
|--|--------------------|------------------------|
| Finalize contract                                | May 2015           | FLDOE, Alpine          |
| Kick off meeting                                 | June               | FLDOE, Alpine, edCount |
| Project plan*                                    | June 19            | Alpine                 |
| Provide technical documentation                  | June – July        | FLDOE, AIR             |
| Review documentation                             | June – July        | Alpine, edCount        |
| Monthly status report*                           | June 30            | Alpine                 |
| Conduct in-person and virtual interviews         | July – August      | Alpine, edCount        |
| Preliminary evaluation reports*                  | July 31            | Alpine, edCount        |
| a. Test development*                             |                    |                        |
| b. Field testing*                                |                    |                        |
| c. Test blueprint & Construction*                |                    |                        |
| d. Test Administration*                          |                    |                        |
| e. Scaling, equating, & scoring*                 |                    |                        |
| f. Specific evaluation of psychometric validity* |                    |                        |
| Monthly status report*                           | July 31            | Alpine                 |
| Final evaluation report*                         | August 28, 2015    | Alpine, edCount        |
| Review draft report                              | August – September | FLDOE                  |
| Monthly status report*                           | September 1, 2015  | Alpine                 |

## Management and Staffing Plan

Alpine and edCount are committed to the successful completion of all aspects of the proposed work. The Work Plan provides specific evaluations will be completed. The proposed team has extensive experience working with large scale clients who require extensive lines of communication and issue tracking. The project will be directed by Dr. Andrew Wiley who has led large scale, nationally recognized educational testing programs. He will be assisted by Dr. Chad Buckendahl and Dr. Ellen Forte for Alpine and edCount, respectively, who are both experienced in conducting large scale assessment system evaluations. In addition, Ms. Erica Brown will serve as the Program Manager for the project, coordinating communications among the organizations, facilitating the sharing of materials and data, and ensuring that project documentation is completed on time and within budget.

As part of the management plan, the project team is committed to delivering all reports in a manner that is consistent with FLDOE expectations. As such, within any report delivered, our team will work to ensure that all conclusions, commendations, and recommendations associated with the project are warranted with sufficient, supporting evidence. Procedurally, prior to submitting any reports to the FLDOE, reports will be reviewed and approved by one or more key staff members and the project director. All reports will also be reviewed by a style editor to ensure that the report is free from spelling and grammatical errors.

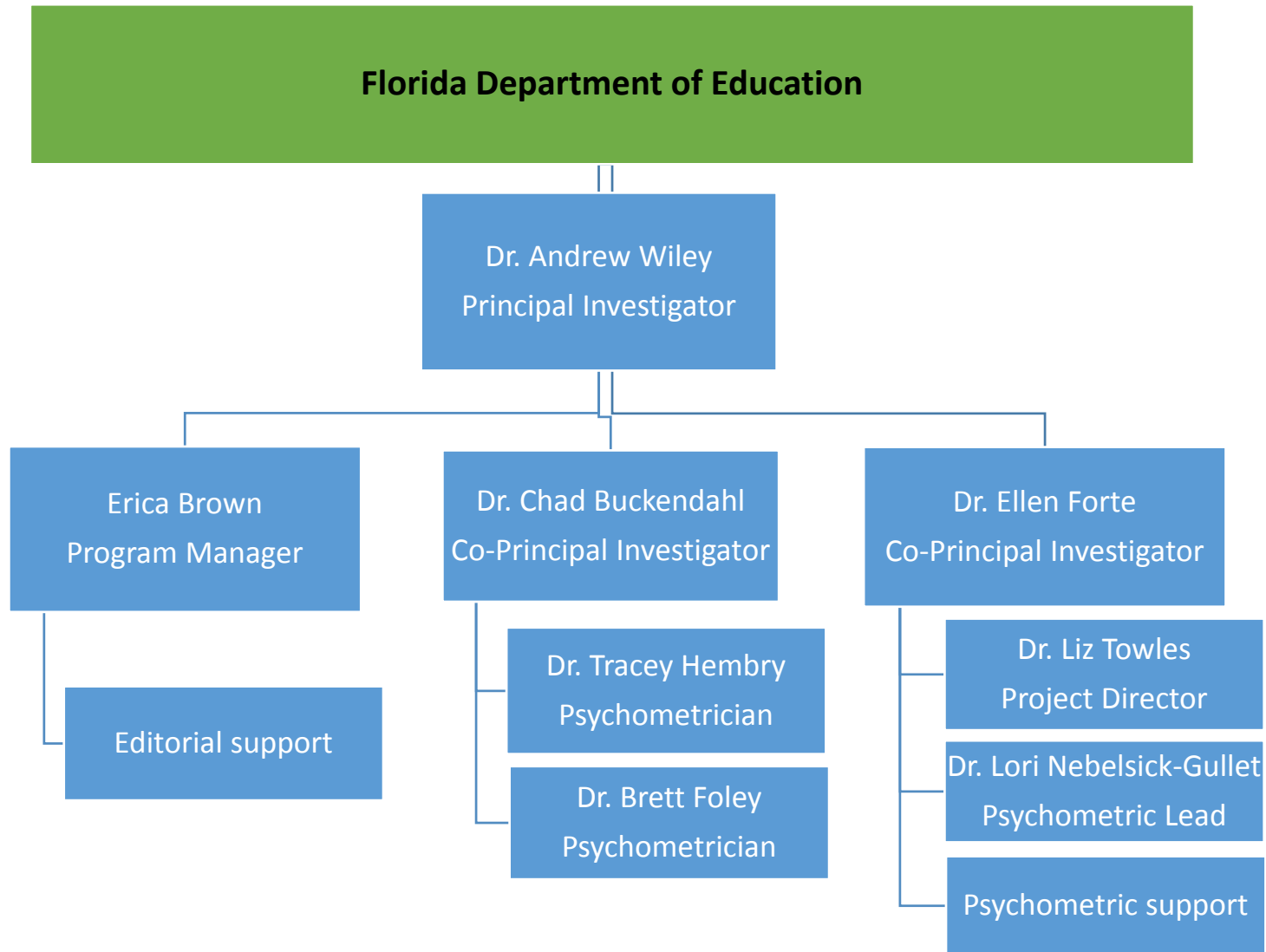
With respect to managing sensitive documentation and data, our team has extensive experience working with a wide variety of data files across a range of constituents. To protect the security and integrity of the data, a number of key principles will be followed in the collection, management, and sharing of all data files:

- During the collection of data files as part of the implementation of the evaluation design, personal identifying information will not be collected from teachers or students. Any personal identifying information will not be recorded and individual records will be masked to ensure that individuals cannot be identified from the data file.
- All individuals working on this project who have access to data files will be trained on the appropriate handling of personal identifying information.
- No data files will be shared over email. All data files will be shared via authorized, secure data transfer methods (e.g., FTP sites, FLDOE-specific Sharepoint site).
- All data files will be stored behind a firewall on a secure file store.

To reinforce our commitment to defending sensitive information, Alpine also maintains Safe Harbor U.S.-E.U. (European Union) certification with respect to the European Commission's Directive on Data Protection.

The following organizational chart illustrates the management structure we will use to successfully fulfill the scope of this project followed by descriptions of key staff members for each of the partner organizations.

Organizational Chart



## Alpine Testing Solutions

*Andrew Wiley, Ph.D., Senior Psychometrician, Director of Education Services.* Dr. Wiley will serve as the Principal Investigator for the project. Dr. Wiley has over 15 years of experience in the education and certification/licensure fields and will serve as the lead psychometrician for the project to guide the design and implementation of the scope. He has worked with high stakes programs in the education field such as the SAT, and with computer based test (CBT)-delivered programs. While at The College Board, Dr. Wiley worked on setting college readiness standards and managed the test development activities as well as the scaling and equating for the SAT. Dr. Wiley is active in the measurement community, and just completed his service on the Board of Directors for the Association of Test Publishers (ATP), as well as Chair of the National Council on Measurement in Education (NCME) Annual Award Committee.

*Chad W. Buckendahl, Ph.D., Senior Psychometrician.* Dr. Buckendahl will serve as Co-Principal Investigator for the project, working directly with Dr. Wiley in the development and design of the project activities. Dr. Buckendahl has worked with a range of student assessment programs in a number of US states (e.g., NC, NE, NV, SD, WA, WY); licensure and professional certification testing programs (e.g., architecture, dentistry, dental hygiene, law, education); and advising on assessment development, validation, and related policy considerations for general, end of course, alternate, collection of evidence, and English language literacy. He also led the most recent *Evaluation of the National Assessment of Educational Progress* for US Department of Education. In addition to providing consultation, he has conducted audits, led validation research, and served as an expert witness on psychometric issues in legal challenges.

*Erica J. Brown, AIA, NCARB, Program Manager II.* Ms. Brown will serve as the Program Manager for the project. She has more than 10 years of experience in the assessment industry, specifically in the licensure and professional certification arena. In her role, she focuses on innovative and practical solutions for clients, organization of complex projects, and financial and resource management. Prior to joining Alpine, Ms. Brown was Director of Examinations, with oversight of the development, administration, and management of a large-scale licensure testing program supported by a multi-vendor solution that mirrors the key vendors listed within this proposal. As a Program Manager, she is responsible for assisting clients with the process of coordinating programmatic evaluations, facilitating communications and handoffs with resources, developing testing programs and ensuring client satisfaction with outcomes.

*Tracey Hembry, Ph.D., Psychometrician II.* Dr. Hembry is a psychometrician with Alpine. She has worked with a range of testing programs, including K-12 state educational assessment and educator licensure. Additionally, she has provided general educational measurement and related policy consultation, stakeholder use of assessment score information, and has led validation research. Dr. Hembry received her Ph.D. from the University of Iowa. Her research interests include teacher effectiveness, growth models, alignment, and standard setting.

*Brett P. Foley, Ph.D., Psychometrician II.* Dr. Foley is a psychometrician with Alpine. He has worked with licensure and certification testing programs in architecture, dentistry, and hearing

devices; and state student assessment programs in Hawaii, Nebraska, Nevada, South Dakota, Washington, and Wyoming. In addition to providing general educational measurement consultation, he has conducted a number of psychometric audits of testing programs, and led research in alignment and standard setting in education and credentialing. Within state assessment programs, these evaluation and research activities have been for general education assessments, alternate assessments, and English language literacy assessments.

### [edCount, LLC](#)

*Ellen Forte, Ph.D., CEO and Chief Scientist.* Dr. Forte will serve as a Co-Principal Investigator for the project, working closely with Dr. Wiley. With two decades' experience conducting research, providing advice and reporting on standards, assessments, and accountability, Dr. Forte is a respected authority on assisting state and local education agencies in the successful interpretation and implementation of education policies. Dr. Forte serves as the Chief Validity Evaluator for the National Center and State Collaborative, funded by the Office of Special Education Programs at the US Department of Education, which is developing an alternate assessment as well as related curricula and professional development modules. In addition, Dr. Forte served as the Principal Investigator for validity evaluation projects for the alternate academic assessments in the District of Columbia, Puerto Rico, and Hawaii, the federally-funded Language Instruction Educational Programs (LIEPs) evaluation, and was the co-principal investigator for the federally-funded project Evaluating the Validity of English Language Proficiency Assessments (EVEA), which brought together the states of Washington, Montana, Oregon, Idaho, and Indiana. She also served as a Senior Advisor for the first national evaluation of Title III implementation. She has been a peer reviewer of statewide standards and assessment systems under both the Improving America's Schools Act of 1994 and the No Child Left Behind Act of 2001 as well as a reviewer of states' graduation rate submissions and state longitudinal data system proposals for funding from the Institute for Education Sciences.

Dr. Forte serves on the Technical Advisory Committees for ACT, Inc., Louisiana, Montana, South Dakota, and Wyoming. She is a member of the editorial boards for Educational Measurement: Issues and Practice, Applied Measurement in Education, the National Council on Measurement in Education newsletter, the NCLB Advisor, and the "No Child Left Behind Alert" published by Eli Research.

*Elizabeth Towles, Ph.D., Managing Associate and Alternate Assessment Specialist.* Dr. Towles has extensive experience in assessment and validity studies, with a special focus in alternate assessment. She has led and assisted with numerous local, regional, and national studies of both general and alternate assessment systems, has served as coordinator and manager of various projects intended to improve, design, or redesign assessment systems, and has played a pivotal role in the development and evaluation of alternate assessment systems around the country. Dr. Towles has served as Project Director in edCount's contracts to evaluate the validity of alternate assessments for students with the most significant cognitive disabilities in Hawaii, the District of Columbia, and Puerto Rico, and currently serves as the Process Evaluator for the National Center and State Collaborative, funded by the Office of Special Education Programs at the US Department of Education, which is developing an alternate assessment as well as related

curricula and professional development modules. In this role, she analyzes project implementation and operation, provides coordination of project management activities, and oversees the external project evaluation. Over the past decade Dr. Towles has authored or co-authored more than a dozen research reports, the majority of which have appeared in peer-reviewed academic journals. She has designed several research instruments that are in wide use, including an index of learner characteristics, and authored several book chapters.

*Lori Nebelsick-Gullett, Ph.D., Principal Associate.* Dr. Nebelsick-Gullett possesses over two decades of experience as an assessment and evaluation expert, educator, and grant evaluator. Dr. Nebelsick-Gullett is the founder of NG Consulting Services, where she provides technical support to school districts, education related companies, and educational organizations in the areas of accountability, continuous quality improvement, data analysis and use of data in decision-making, quality assessment practices, evaluation, and the application of measurement concepts to assessment and evaluation processes. Under this role, Dr. Nebelsick-Gullett provided program design and evaluation support to numerous projects including the College Board, Microsoft, and EPIC. She also designed and implemented professional learning communities (PLCs) sessions for teachers, and provided training for teachers and administrators on formative assessment and descriptive feedback, assessment literacy, data use in the classroom, and alignment of essential learning targets with assessments and state standards. Further, Dr. Nebelsick-Gullett has evaluated numerous grants funded by the U.S. Department of Education and the Colorado Department of Education, and has taught educational measurement, statistics, and research methods at various institutions including the University of Missouri, Baker University, the University of Nebraska Medical Center, and the University of Nebraska-Lincoln. She served as the lead reviewer for interim assessment items for a College Board project, was a member of the Peer Review team for the Nebraska Department of Education, a member of the Nebraska District Assessment Evaluation Team, and also served on the Advisory Council for the development of a web-based assessment system aligned with the Texas Essential Knowledge and Skills.

Brief resumes for key staff members associated with the project are included as Appendix A.

## Summary

Alpine and edCount appreciate the opportunity to collaborate with the Florida Department of Education and to assist with conducting this independent verification of the psychometric validity for the Florida Standards Assessment program. Our staff members' collective experience working with organizations that are required to balance the often competing needs of psychometrics, policy, practical, and business realities serves as a unique benefit to the Department of Education and the State of Florida. Anchored by the industry standards, the evaluation design that we have proposed is also flexible in terms of the structure and requests for additional emphasis of different elements of test development and validation that may emerge during the evaluation inquiry. Thank you for your consideration.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Buckendahl, C. (April, 2015). *Accreditation of credentialing program: An analogous model for educational assessments?* Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL.
- Buckendahl, C., Plake, B., Davis, S., Sireci, S., Hambleton, R., Zenisky, A., Wells, C., & Keller, L. (2009). *Evaluation of the National Assessment of Educational Progress: Final Report*. Washington, D.C.: U.S. Department of Education
- Buckendahl, C. & Plake, B. (2006). Evaluating tests. In S. Downing & T. Haladyna (eds.) *Handbook of Test Development* (pp. 725-738). Mahwah, NJ: Erlbaum.
- Davis-Becker, S. & Buckendahl, C. (2013). A proposed framework for evaluating alignment studies. *Educational Measurement: Issues and Practice*, 31(1), 23-33.
- Forte, E. (April, 2015). *Peer review in policy and practice*. Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL.
- National Council on Measurement in Education (2012). *Testing and data integrity in the administration of statewide student assessment programs*. Madison, WI: Author.
- Webb, N. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (Research Monograph No. 6). Washington, DC: Council of Chief State School Officers.
- Wiley, A. (April, 2015). *Policies and procedures for the independent evaluation of educational assessment programs*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.



## Appendix A: Resumes of Key Staff

## ANDREW WILEY

---

### *Education*

- 1999 Ph.D. Psychometrics and Quantitative Psychology  
Fordham University, Bronx NY
- 1992 M.A. Psychometrics and Quantitative Psychology  
Fordham University, Bronx NY
- 1991 B.A. Psychology (minor in English)  
LaSalle University, Philadelphia PA

### *Experience*

- 2013 – Current Senior Psychometrician, Director of Education Services  
Alpine Testing Solutions, Inc.
- 2002 – 2013 Associate Research Scientist, Senior Director, Executive Director  
The College Board
- 1999 – 2002 Psychometrician  
American Board of Internal Medicine
- 1996 – 1999 Psychometrician  
American Institute of Certified Public Accountants
- 1995 – 1996 Research Associate  
Association of American Medical Colleges

### *Selected Scholarly Research*

- Wiley, A. (April, 2015). *Policies and procedures for the independent evaluation of educational assessment programs*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Wiley, A. & Davis-Becker, S. (April, 2015). *Developing a framework for the international benchmarking of performance standards*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Merriman, J., Sireci, S., Wiley, A., & Farrar Hamen, C. (April, 2014). *A validity framework for a scientific knowledge for teaching assessment*. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA.
- Camara, W., Packman, S., & Wiley, A. (2013). College, graduate and professional school admissions testing. In K.F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology, Volume 3: Testing and assessment in school psychology and education*. American Psychological Association.
- Gierl, M. & Wiley, A. (September, 2012). *Policies and procedures for the development of more effective item writing and test development*. Paper presented at the annual meeting of the Association of Test Publishers - Europe, Berlin, Germany.

- Wiley, A., Douglas, A., Fiumana, D., Johnson, M., & Kuan, D. (April, 2012). *Development of a taxonomy of attributes and components of test items: The foundation of a more efficient assessment design process*. Paper presented at the annual meeting of the National Council on Measurement in Education. Vancouver, BC.
- Wyatt, J. N., Kobrin, J., Wiley, A., Camara, W. J., & Proestler, N. (2011). *SAT Benchmarks: Development of a College Readiness Benchmark and its Relationship to Secondary and Postsecondary School Performance* (College Board Research Report No. 2011-5). New York, NY: College Board.
- Antal, J., Melican, G., Proctor, T., & Wiley, A. (May, 2010). *The effect of anchor test construction on scale drift*. Paper presented at the annual meeting of the National Council on Measurement in Education. Denver, CO.
- Kaliski, P. K., Antal, J., & Wiley, A. (May, 2010). *Examining individual and school effects on the revised SAT scores using multilevel modeling*. Paper presented at the annual meeting of the American Educational Research Association. Denver, CO.
- Padaki, M., Harris, W.G., Dhiremdra, R., Bhardwaj, M., & Wiley, A. (March, 2010). *Enhancing the effectiveness of examinations by leveraging technology*. Paper presented at the Emerging Directions in Global Education conference, New Delhi, India.
- Wiley, A., Wyatt, J. N., & Camara, W. J. (2010). *The Development of a Multidimensional College Readiness Index* (College Board Research Report No. 2011-5). New York, NY: College Board.
- Wyatt, J. N. & Wiley, A. (May, 2010). *Academic rigor: Differences in high school course difficulty among ethnic subgroups and its implications for college performance*. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.
- Camara, W., Melican, G., & Wiley, A. (February, 2009). *Developing a new test from conceptualization to maintenance*. Paper presented at the annual meeting of the Association of Test Publishers, Palm Springs, CA.

## CHAD W. BUCKENDAHL

---

### *Education*

- 2000 Ph.D. Quantitative and Qualitative Methods in Education, University of Nebraska, Lincoln
- 1996 M.L.S. Legal Studies, University of Nebraska College of Law
- 1994 B.A. Political Science, University of Nebraska, Lincoln

### *Professional Experience*

- 2007 – pres. Senior Psychometrician and Director of Strategic Partnerships, Alpine Testing Solutions.
- 1998 – 2007 Director (2002-2007), Assistant Director (2000-2002), Research Associate (1998-2000), Buros Institute for Assessment Consultation and Outreach, University of Nebraska, Lincoln.
- 1990 – 1997 Computer Programmer (1994-1997), Survey Researcher (1990-1994), The Gallup Organization, Inc.

### *Selected Scholarly Research*

- Buckendahl, C. (April, 2015). *Accreditation of credentialing program: An analogous model for educational assessments?* Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL.
- Buckendahl, C. & Foley, B. (April, 2015). *Policy linking as cut score moderation: Considerations for practice.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Buckendahl, C. (April, 2013). *A practitioner's guide to validation framework development.* Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA.
- Davis-Becker, S. L. & Buckendahl, C. W. (2013). A proposed framework for evaluating alignment studies. *Educational Measurement: Issues and Practice*, 31(1), 23-33.
- Ferdous, A. A. & Buckendahl, C. W. (2013). Evaluating panelists' standard setting perceptions in a developing nation. *International Journal of Testing*, 13(1), 4-18.
- Davis-Becker, S. L. & Buckendahl, C. W. (2013). Identifying and evaluating external validity evidence for passing scores. *International Journal of Testing*, 13(1), 50-64.
- Lim, G. S., Geranpayeh, A., Khalifa, H., & Buckendahl, C. W. (2013). Standard setting to an international reference framework: Implications for theory and practice. *International Journal of Testing*, 13(1), 32-49.
- Buckendahl, C. & Davis-Becker, S. (2012). Setting passing standards for credentialing examinations. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2<sup>nd</sup> ed., pp. 485-502). Routledge.

- Lim, G., Buckendahl, C., & Foley, B. (July, 2012). *Evaluating test design through complementary standard setting studies*. Paper presented at the International Test Commission conference. Amsterdam, The Netherlands.
- Davis-Becker, S., Buckendahl, C., & Gerrow, J. (2011). Implications of random ordering on the bookmark standard setting method. *International Journal of Testing, 11*(1), 24-37.
- Ferdous, A., Bechard, S., & Buckendahl, C. (2011). Setting performance standards on alternate assessments for students with disabilities. In J. Bovaird, K. Geisinger, & C. Buckendahl (Eds.), *High stakes testing in education: Science and practice in K-12 setting* (pp. 43-58). Washington, DC: American Psychological Association.
- Morgan, D. & Buckendahl, C. (April, 2011). *Evaluating a strategy for setting cut scores on a computer adaptive test*. Paper presented at the annual meeting of the National Council on Measurement in Education. New Orleans, LA.
- Buckendahl, C., Ferdous, A. & Gerrow, J. (2010). Recommending cut scores with a subset of items: An empirical illustration. *Practical Assessment, Research & Evaluation, 15*(6). Available online: <http://pareonline.net/getvn.asp?v=15&n=6>.
- Buckendahl, C., Davis, S., Plake, B., Sireci, S., Hambleton, R., Zenisky, A. & Wells, C. (2009). *Evaluation of the National Assessment of Educational Progress: Final Report*. Washington, DC: U.S. Department of Education
- Buckendahl, C. W., Plake, B. S., & Davis, S. L. (2009). Conducting a lifecycle audit of the National Assessment of Educational Progress. *Applied Measurement in Education, 22*(4), 321-338.
- Davis, S., Buckendahl, C. & Plake, B. (2008). When adaptation is not an option: An application of multilingual standard setting. *Journal of Educational Measurement, 45*(3), 287-304.
- Norman, R. & Buckendahl, C. (2008). Determining sufficient measurement opportunities when using multiple cut scores. *Educational Measurement: Issues and Practice, 27*(1), 37-46.
- Buckendahl, C. & Plake, B. (2006). Evaluating Tests. In S. Downing & T. Haladyna (Eds.), *Handbook of Test Development* (pp.725-738). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Buckendahl, C., Huynh, H., Siskind, T., & Saunders, J. (2005). A case study of vertically-moderated standard setting for a state science assessment program. *Applied Measurement in Education, 18*(1), 83-98.
- Buckendahl, C., Plake, B. & Impara, J. (2004). A strategy for evaluating district developed assessments for state accountability. *Educational Measurement: Issues and Practices, 23*(2), 17-25.
- Buckendahl, C., Smith, R., Impara, J. & Plake, B. (2002). A comparison of Angoff and Bookmark standard setting methods. *Journal of Educational Measurement 39*(4), 253-263.

# Erica J. Brown

---

## *Education*

- 1998                    **B.S.** – Architecture  
Ball State University, Muncie, IN
- 1998                    **B.S.** – Environmental Design  
Ball State University, Muncie, IN

## *Professional Experience*

- 2013 – Current        **Program Manager**  
Alpine Testing Solutions  
Raleigh, NC
- Develop and maintain strong business relationships with clientele. Oversee education, certification and licensure projects (e.g., test development, validation studies) to meet business and client objectives. Allocate resources and monitor funding to stay within scoped proposals. Manage internal and external resources to ensure deadlines are met. Create and supervise schedules to assure that deliverables are completed and on time within professional standards.
- 2003 – 2013            **Director, Examination (2007-2013), Assistant Director, ARE Development (2003-2007)**  
National Council of Architectural Registration Boards (NCARB)  
Washington, D.C.
- Progressive responsibilities throughout tenure with the organization. Ultimately directed the national examination program for architect licensure. Directed eight member staff facilitating 6-8 committees of volunteers. Oversaw 33,000-36,000 administrations per year across seven examinations in the program. Managed program activities for 130-160 content development volunteers annually. Maintained contract obligations and oversight of exam vendor contracts.
- 1998 – 2003            **Project Architect (2003), Intern Architect (1998-2002)**  
InterDesign  
Indianapolis, IN
- Oversaw construction projects from concept to completion, managing internal and external team members. Frequently interfaced with engineers and contractors to deliver projects on time and within budget requiring minimal change orders. Delivered executive presentations with design options. Actively engaged in review of submittals, payment applications, and ensured compliance with contract requirements.

***Professional Credentials and Memberships***

**American Institute of Architects (AIA)**

**American Society of Association Executives (ASAE)**

**Certified Scrum Master – Scrum Alliance**

**Construction Document Technologist – Construction Specifications Institute**

**NCARB Certified - NCARB**

***Selected Professional Conference Presentations***

**2012** Institute for Credentialing Excellence

**2011** Performance Testing Council

**2008** Association of Test Publishers



## **TRACEY HEMBRY (MAGDA)**

---

### ***Education***

2009                      Ph.D. Educational Measurement and Statistics  
The University of Iowa, Iowa City, IA

2006                      M.A. Educational Measurement and Statistics  
The University of Iowa, Iowa City, IA

2004                      B.S. Psychology (minor in Statistics)  
James Madison University, Harrisonburg, VA

### ***Experience***

2015 – Current        Psychometrician II  
Alpine Testing Solutions

2009 – 2015         Senior Research Scientist, Manager, Psychometrician  
Pearson

2007 – 2009         Research Assistant  
Institute of Education Sciences (IES) Unsolicited Research Grant

2006 – 2007         Research Assistant  
National Assessment of Educational Progress (NAEP) Secondary Analysis  
Grant

2004 – 2006         Research Assistant  
Iowa Testing Programs

### ***Service***

National Council on Measurement in Education (NCME)

    Paper reviewer for annual meeting

    Chair, Graduate Student Issues Committee (GSIC)

American Educational Research Association (AERA)

    Paper reviewer for annual meeting

### ***Selected Scholarly Research and Presentations***

- Hembry, T.R., Bradshaw, L.P., Kobrin, J.L., & Sovde, D. (2015, June). Meaningful reporting of Diagnostic Classification Model (DCM) data. Accepted for presentation at the annual National Conference on Student Assessment, San Diego, CA.
- Wei, H. Hembry, T., Murphy, D.L., McBride, Y. (2012, May). Value-added models in the evaluation of teacher effectiveness: A comparison of models and outcomes. *Pearson Research Report*.
- Hembry, T.R. (2012, February) *A practical guide to developing educator effectiveness systems*. Presented at the annual meeting of the California Association of African-American Superintendents and Administrators, Sacramento, CA.
- Magda, T. (2011, March 14). Teacher effectiveness measures: The tortoise and the hare. *Pearson's TrueScores blog*. <http://www.truescores.com/2011/03/teacher-effectiveness-measures-tortoise.html>
- O'Malley, K. McClarty, K., Magda, T. & Burling, K. (2011, October). *Making sense of the metrics: Student growth, value-added models, and teacher effectiveness*. Paper presented at the 42<sup>nd</sup> annual meeting of the Northeastern Educational Research Association, Hartford, CT.
- Magda, T. R. & Ho, A. D. *The Impact of Recent Changes to NCLB: How Growth Models and Index Systems Alter Trends Below Proficiency*. Presented at the 2009 Annual Meeting of the American Educational Research Association in San Diego, CA.
- Ho, A. D. & Magda, T. R. (2008, March). *Growth Toward Which Standard?: The Dependency of Growth Models on Proficiency Standards*. Presented at the 2008 Annual Meeting of the American Educational Research Association in New York City.
- Magda, T. R. & Ho, A.D. (2008, March). *Comparing Trends at Different Cut Scores: Distribution-Wide NCLB and NAEP Comparisons from 2003 to 2007*. Presented at the 2008 Annual Meeting of the American Educational Research Association in New York City.
- Thiessen, B. A., Magda, T. R. & Ho, A. D. (2008, March). *Nonparametric Comparisons of High-Stakes and Low-Stakes Trends and Gaps: 2003-2007*. Presented at the 2008 Annual Meeting of the American Educational Research Association in New York City.
- Magda, T. R. (2007, April). *Implementation Recommendations for Use of Text-Reader Software as an Assessment Accommodation*. Presented at the 2007 Annual Meeting of the American Educational Research Association in Chicago, Illinois.
- Magda, T. R. & Ho, A. D. (2007, April). *A Distribution-Wide View of High-Stakes Test Score Trends*. Presented at the 2007 Annual Meeting of the American Educational Research Association in Chicago, Illinois.

# Brett P. Foley

---

## ***Education***

|      |  |
|------|--|
| 2010 | Ph.D. – Quantitative, Qualitative, and Psychometric Methods<br>University of Nebraska, Lincoln, NE |
| 2003 | M.S. Biometry (minor in Survey Research and Methodology)<br>University of Nebraska, Lincoln, NE    |
| 2001 | B.S. – Mathematics (minor in Computer Science)<br>Chadron State College, Chadron, NE               |

## ***Professional Experience***

|                |  |
|----------------|--|
| 2010 – Current | Psychometrician II, Psychometrician I<br>Alpine Testing Solutions, Inc.  |
| 2005 – 2010    | Interim Asst. Director, Project Coordinator, Graduate Assistant<br>Buros Institute for Assessment Consultation and Outreach<br>University of Nebraska, Lincoln |
| 2001-2006      | Instructor, Graduate Teaching Assistant<br>University of Nebraska, Lincoln   |
| 2005-2006      | Statistics and Measurement Consultant<br>Nebraska Center for Research on Children, Youth, Families and Schools   |
| 2004-2006      | Statistical Consultant<br>Nebraska Evaluation and Research Center  |
| 2002-2005      | Statistical Analyst/Data Analyst Intern<br>Nebraska Health and Human Services System, Dept. of Epidemiology  |

## ***Service***

- Northern Rocky Mountain Educational Research Association (Executive Board member 2007-present; President 2011-2012)
- National Council on Measurement in Education – Website Committee Member
- Council of Chief State School Officers, English Language Development Assessment State Collaborative on Assessment and Student Standards (TAC Member, 2010)
- Reviewer – *Applied Measurement in Education, Educational Assessment*

## ***Scholarly Research***

### Selected Recent Publications

- Foley, B.P., & Buckendahl, C.W. (2013). Using visual displays to inform assessment design and development. In Gregg Schraw, Matt McCrudden, & Dan Robinson (Editors). *Learning through visual displays*. Charlotte, NC: Information Age Publishing.
- Foley, B. P., Dwyer, A. C., Chuah, D., Rawls, A. (Producers), & Foley, B. P. (Director). (2013). *Testing in the movies and on television* [Motion Picture]. United States: National Council on Measurement in Education.
- Buckendahl, C.W. & Foley, B.F. (2011). High Stakes Uses of Intelligence Testing. In James Bovaird, Kurt Geisinger, & Chad Buckendahl (Editors). *High Stakes Testing in Education - Science and Practice in K-12 Settings [Festschrift to Barbara Plake]*. American Psychological Association Press, Washington, DC.
- Doll, B., Spies, R.A., LeClair, C., Kurien, S., & Foley, B.P. (2010). Student perceptions of classroom learning environments: Development of the ClassMaps Survey. *School Psychology Review*, 39, 203-218.
- Foley, B.P. (2008, April 2). Tax breaks based on test scores could boost parental involvement. *Omaha World Herald*, p. 7B

### Selected Recent Presentations

- Lim, G. S., Buckendahl, C. W., & Foley, B.P. (2012, July). *Evaluating test design through complementary standard setting studies*. Paper presented at The 8th Conference of the International Test Commission, Amsterdam, The Netherlands.
- Foley, B.P., & Lupher, D.A.. (2011, October). *Evaluating the stability of cut scores over time in a statewide assessment system*. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC.
- Foley, B.P. (2011, October). *Validity by design: A validity-centered assessment development approach*. Workshop presented at the Annual meeting of the Northern Rocky Mountain Educational Research Association, Jackson Hole, WY.
- Foley, B. P. (2011, April). *Realistic Expectations: State-level Changes in the Percentage of Proficient Students 2002-2008*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Dwyer, A.C., & Foley, B. F. (2010, May). *Invariance of IRT equating across subpopulations of different abilities and year-to-year growth rates*. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.
- Foley, B.P., Geisinger, K.F., Roschewski, P., & Foy, E. (2009, October). *Conducting an alignment study in the context of a performance assessment with a single writing prompt*. Paper presented at the Annual meeting of the Northern Rocky Mountain Educational Research Association, Jackson Hole, WY.
- Foley, B. F. (2009, April). *Improving IRT item parameter estimates with small sample sizes: Evaluating the efficacy of a new data augmentation technique*. Paper presented at the annual meeting of the National Council on Measurement in Education. San Diego, CA.
- Chin, T., Rodeck, E., Buckendahl, C.W., & Foley, B. F. (April, 2008). *The Impact of Consensus on Alignment Judgments*. Paper presented at the annual meeting of the National Council on Measurement in Education. New York, NY.

## **Ellen Forte**

---

### ***Education***

***Ph.D. Educational Psychology, 1996***

University of Iowa

***M.A. Educational Psychology, 1994***

University of Iowa

***B.A. Physical Education & Dance, 1987***

University of Iowa

### ***Present Position***

**CEO & Chief Scientist, edCount, LLC**

*2003 – present*

Founder edCount, LLC, a professional services firm specializing in education assessment, evaluation, data management, reporting, and accountability. Major projects and clients include:

**National Centers and State Collaborative Alternate Assessment Project** (September 2010 – 2015)

**Puerto Rico Policy and Technical Assistance Project** (January 2010 – 2014) – Served as the Principal Investigator for a comprehensive system of supports for the Puerto Rico Department of Education that encompasses validity studies.

**Evaluating the Validity of English Language Proficiency Assessments** (October 2009 – September 2011)

**State Departments of Education, State Boards of Education, and Legislative and Appointed Taskforces** (multiple states; 2003 to present)

**National Alternate Assessment Center** (September 2007 – September 2011)

**US Department of Education** (1998 – present) –Currently involved in reviews of several state standards and assessment systems.

**Council of Chief State School Officers** (1999 – 2008) – Consulted on assessment and accountability issues with representatives of state departments of education from across the country.

### ***Prior Professional Experience***

**Director of Student Assessment, Baltimore City Public Schools 2002-2003**

**Senior Research Analyst, American Institutes for Research 2000-2002**

**Education Consultant, Bureau of Student Assessment and Research, Connecticut State Department of Education 1997-2000**

**Project Director, National Evaluation Systems, Inc. 1996-1997**

### ***Books, Journal Articles, Book Chapters, & Monographs***

- Perie, M. & Forte, E. (2011). *Developing a validity argument for assessments of students in the margins*. In M. Russell (Ed). *Assessing Students in the Margins*. Information Age Publishing.
- Forte Fast, E. & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington DC: Council of Chief State School Officers.
- Forte Fast, E. (2002) *A Guide to Effective Accountability Reporting*. Washington DC: Council of Chief State School Officers and US Department of Education.

### ***Keynotes and Invited Presentations***

- Forte, E. (April, 2014). *Argument in action: Implementing validity theory in the real world*. Invited presentation for the AERA Special Interest Groups (SIG) on Professional Licensure and Certification; Test Validity Research and Evaluation; and Cognition and Assessment at the Annual Meeting of the American Educational Research Association, Philadelphia, PA.
- Forte, E. (June, 2013). *2013 Assessment Bootcamp – An Overview of the Practical and Technical Issues in Large-Scale Assessment Programs*. (Invited organizer and presenter.) Session presented at the Annual National Conference on Student Assessment, Washington, DC.
- Forte, E. (November, 2007). *Assessment and accountability post-NCLB: The next generation*. Invited Keynote Address for the Annual Conference of the California Educational Research Association, Dana Point, CA.
- Forte Fast, E. (April, 2005). *Examining validity issues in state accountability systems*. Organizer and co-facilitator (with Brian Gong and Arthur Coleman) for a training pre-session at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada.
- Forte, E. & Perie, M. (November, 2013). *Considering instructional sensitivity in the validity evaluation process*. Paper presented at the first annual Instructional Sensitivity Conference, Lawrence, KS.
- Forte, E. (June, 2013). *Transitioning assessment systems*. Paper presented at the Annual National Conference on Student Assessment, Washington, DC.
- Forte, E. (October 2012). *A next generation of validity evaluation and technical documentation*. Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Rocky Hill, CT.
- Forte, E. (June, 2009). *Validity evaluation of alternate achievement standards assessments: Emerging findings and building a validity argument*. Paper presented at the Council of Chief State School Officers' National Conference on Student Assessment, Los Angeles, CA.

## **Lori Nebelsick-Gullett**

---

### ***Education***

#### ***Ph.D. Educational Psychology***

University of Nebraska, Lincoln

#### ***M.S. Exercise Physiology***

University of Nebraska, Lincoln

#### ***B.A. Interdisciplinary Science***

Morningside College

### ***Present Position***

#### **Principal Associate, edCount, LLC**

*2013 – present*

Provide substantive and managerial leadership on large-scale client contracts and tasks. Design and manage research activities; pursue and procure contract and grant funding. Manage clients, staff, budgets, and timelines to ensure that clients receive high-quality deliverables by specified delivery dates. Current responsibilities include:

#### **National Center and State Collaborative General Supervision Enhancement Grant –**

Provide leadership and oversight for project-wide technical documentation, ensuring collection of evidence needed to address the project’s Validity Argument and the claims in the Theory of Action for the NCSC AA-AAS; provide support and expertise to development and implementation of research protocols; ensure quality development and implementation of evaluative tools and processes; and support coordination of project management activities.

### ***Professional Experience***

#### **Founder, NG Consulting Services**

*2002 – 2013*

Consultant to school districts, education related companies, and other educational organizations in the areas of accountability, continuous quality improvement, data analysis and use of data in decision making, quality assessment practices, evaluation, and the application of measurement concepts to assessment and evaluation processes. Consultation projects include:

#### **Assessment and Evaluation Specialist, Educational Service Unit 18, Lincoln Public Schools**

*2004 – 2013*

#### **Executive Director, Accountability and Continuous Improvement, Richardson Independent School District**

*1999 – 2002*

#### **Director, Student Performance and Program Evaluation, Richardson Independent School District**

*1997*



**Assessment Professional, Assessment Office, Kansas City Missouri School District**  
*1995 – 1997*

**Program Associate, Test Development and Research, Center for Educational Testing and Evaluation**  
*1993 – 1995*

***Professional Affiliations & Organizations***

Chair of Classroom Assessment SIG; AERA special interest group (2011-12)

Chair of Training and Professional Development Committee, National Council on Measurement in Education (2006); committee member (2007; 2008)

Member of District Assessment Evaluation Team (State of Nebraska)

Trainer of Trainers: Assessment Training Institute

***Selected Publications & Presentations***

Nebelsick-Gullett, L. J., Herrell, A. & Bley, S. (2004) Assessment to Enhance Learning: Instructor Manual and Materials.

Nebelsick-Gullett, L. J. & Asp, E. (2004) Data Analysis and Management for Leadership Teams: Instructor Manual and Materials.

Buckendahl, C. W., Nebelsick-Gullett, L. J., Bandalos, D., Benson, J., & Irwin, P. (2003). Report to the South Carolina Department of Education from Standard Setting Workshops for the Palmetto Achievement Challenge Tests and End-of-Course Examination Program.

## Elizabeth Towles

---

### *Education*

***Ph.D. Educational Psychology, 2007***

University of Kentucky

***Developmental Disabilities Certificate, 2003***

University of Kentucky

***M.S. Education, 2000***

University of Kentucky

***B.A. Art and Psychology, summa cum laude, 1999***

University of Kentucky

### *Present Position*

**Managing Associate and Alternate Assessment Specialist, edCount, LLC 2009 – present**

**National Center and State Collaborative (NCSC) General Supervision Enhancement Grant** – Serve as Process Evaluator to document and analyze project implementation and operations, provide coordination of project management activities, and oversee the external project evaluation. The collaborative consists of 5 partner organizations, 18 states, and the six Pacific Rim entities.

**Evaluating the Validity of English Language Proficiency Assessments (EVEA)** – Served as Project Director, providing oversight and coordination of project management meetings, leadership for development of project instruments, management of grants and contracts, and oversight of administration and partnerships throughout project duration.

### *Professional Experience*

**Research Coordinator, National Alternate Assessment Center Validity General Supervision Enhancement Grant, Lexington, Kentucky 2007 – 2009**

Four overarching goals guided an iterative process of stakeholder involvement and expert review in the validity evaluations for five states' alternate assessments based on alternate achievement standards. The primary goals of this GSEG Consortium were to: 1) demonstrate high quality validity evaluation models through our partnerships with states, 2) provide models of validity-based technical documentation for AA-AAS and eventually for AA-MAS and general assessments, 3) add to the growing research base on high quality, technically sound AA-AAS to provide technical assistance to states as they endeavor to conduct their validity evaluation studies, and 4) provide a range of research-to-practice products that explicate the process and results. Responsibilities included coordination of research activities project.

**Evaluator, National Center for Educational Outcomes General Supervision Enhancement Grant, Lexington, Kentucky 2007 – 2009**

This project sought to develop high-quality assessment and instructional systems that improve not only the accountability system, but also the learning of those students whose disability has precluded them from achieving grade-level proficiency and whose progress is such that they are not likely to reach grade-level proficiency in the same time frame as other students. The project

served to: (a) develop clear and appropriate guidelines for individualized education program (IEP) teams to use in determining which students should be assessed; and (b) develop State assessments using universal design principles based on modified academic achievement standards. Responsibilities included evaluation of the research activities across the project.

### ***Selected Publications & Presentations***

- Kleinert, H., Towles-Reeves, E., Quenemoen, R., Thurlow, M., Fluegge, L., Weseman, L., & Kerbel, A. (2015). Where Students with the Most Significant Cognitive Disabilities Are Taught: Implications for General Curriculum Access. *Exceptional Children, 2*(29).
- Musson, J., Thomas, M., Towles-Reeves, E., & Kearns, J. (2010). An analysis of state alternate assessment participation guidelines. *Journal of Special Education, 44*(2), 67-78.
- Towles-Reeves, E., Kleinert, H., & Muhomba, M. (2009). Alternate assessment: Have we learned anything new? *Exceptional Children, 72*(2), 233-252.
- Towles-Reeves, E., Garrett, B., Burdette, P., & Burdge, M. (2006). What are the consequences? Validation of large-scale alternate assessment systems and their influence on instruction. *Assessment for Effective Intervention, 31*(3), 45-57.
- Towles-Reeves, E., & Kleinert, H. (2006). The impact of one state's alternate assessment upon instruction and IEP development. *Rural Special Education Quarterly, 25*(3), 31-39.

### ***Selected Book Chapters***

- Towles-Reeves, E. (2008). Alternate assessment: Leaving no child behind amidst standards based reform. In R. Mitchell & R. Johnson (Eds.), *Testing Deaf Students in an Age of Accountability*. Washington, D.C.: Gallaudet University Press.
- Kearns, J., Towles-Reeves, E., Kleinert, H., & Kleinert, J. (2009). Who are the children who take alternate achievement standards assessments? In B. Schafer & B. Lissitz (Eds.), *Alternate Assessments based on Alternate Achievement Standards: Policy, Practice, and Potential*. Baltimore, MD: Brookes Publishing.
- Kleinert, H., & Towles-Reeves, E. (2010). What we have learned from alternate assessment research and what we still need to know. In J. Kearns & H. Kleinert (Eds.), *Meaningful Outcomes for Students with Significant Cognitive Disabilities: Alternate Assessments on Alternate Achievement Standards*. Brookes Publishing.

### **National Conference Presentations:**

- Flowers, C., Turner, C., Towles-Reeves, E., Thurlow, M., Davidson, A., & Hagge, S. (2015). *Developing a large-scale assessment using components of evidence-centered design: Did it work?* Paper presented at the annual conference of the National Council on Measurement in Education. Chicago, IL.
- Nebelsick-Gullett, L., Towles-Reeves, E., Perkins, A., & Deters, L. (2015). *Evaluating the quality and impact of items, products, and procedures: NCSC writing alternate assessment based on alternate achievement standards*. Paper presented at the annual conference for the American Educational Research Association. Chicago, IL.
- Kearns J., Thurlow M., Towles-Reeves, E. (2009). *Who are the Students in Alternate and Modified Achievement Standards Assessments?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

## Attachment 1 – Price Sheet

**ATTACHMENT 1**

**PRICE SHEET  
INDEPENDENT VERIFICATION OF THE PSYCHOMETRIC VALIDITY FOR  
THE FLORIDA STANDARDS ASSESSMENT**

We propose to provide the services specified in the Request for Offers. All work shall be performed in accordance with this Request, which has been reviewed and understood. The below price is all inclusive. There shall be no additional costs charged for work performed under Request.

| <b>DESCRIPTION</b>   | <b>TOTAL COST</b>      |
|--|------------------------|
| Independent verification of the psychometric validity for the Florida Standards Assessment | \$ <u>594,310.43</u> * |

\*AWARD MADE BASED ON THIS PRICE

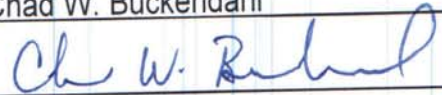
**SIGN BELOW. UNSIGNED OFFERS WILL NOT BE CONSIDERED.**

**VENDOR NAME:** Alpine Testing Solutions, Inc.

**MAILING ADDRESS:** 51 W. Center St. #514

**CITY/STATE/ZIP:** Orem, UT 84057

**AUTHORIZED AGENT (typed):** Chad W. Buckendahl

**AUTHORIZED AGENT (manual):** 

**DATE:** May 26, 2015 **TELEPHONE: (O):** 702-586-7386 **(C):** 402-770-0085 **FAX:** 702-586-7386

**E-MAIL ADDRESS:** chad.buckendahl@alpinetesting.com

## Attachment 2 – Disclosure Statement

# ATTACHMENT 2

## DISCLOSURE STATEMENT

---

### PARTNERSHIP OR INDIVIDUAL

I hereby certify that I, if an individual, or each of us, if a partnership, doing business as \_\_\_\_\_  
(Name of Individual or Partnership)

(am)(is) not now involved in nor have I ever engaged in any private business venture or enterprise, directly or indirectly, with the Commissioner of Education, the Deputy Commissioner of Education, any Associate Commissioner of Education, Division Director, or Bureau Chief within the Florida Department of Education.

I further certify that neither I, nor any partner, if a partnership, nor anyone acting in my or our behalf has requested that any of the above designated persons or any other employee of the Department of Education exert any influence to secure the appointment of \_\_\_\_\_ under this proposed agreement.  
(Name of Individual or Partnership)

(1) \_\_\_\_\_  
Signature

\_\_\_\_\_  
Signature

(1) If partnership, each partner must sign and execute. Signature

---

### COMPANY OR CORPORATION

I hereby certify that neither I nor any owner, officer, director, or shareholder

Alpine Testny Solutions  
(Name of Corporation/Company)



Utah

(1) corporation, licensed to do business in Florida, is presently involved

(Name of State of Inc.)

in or has been engaged in any private business venture or enterprise, directly, or indirectly, with the Commissioner of Education, the Deputy Commissioner of Education, any Associate Commissioner of Education, Division Director, or Bureau Chief within the Florida Department of Education.

I further certify that neither I nor any owner, officer, director, or shareholder of this corporation or anyone acting on behalf of this corporation or any of its owners, officers, directors, or shareholders has requested that any of the above designated persons or any other employee of the Department of Education exert any influence to secure the appointment of

Alpine Testing Solutions, Inc.

under this proposed agreement.

(Company)

(Corporation)

Ch W. Bulbul

(2) Signature

Director of Strategic Partnerships

Title

(1) If company is not incorporated, insert "not incorporated" in this space.

(2) If incorporated, this statement is to be executed by same person who will execute contract, if awarded.