Ohio | Department of Education

ODE Competitive Grants Reader Training and Scoring Process

Issue Date: March 11, 2009 **Revised:** March 10, 2009
Source: Federal Programs
Keywords: Competitive grants; DER, district efficiency rating; grant review; monitoring; Enhancing education Through Technology; 21st CCLC, 21st Century Community Learning Centers; Even Start; Adult Basic Literacy Education; ABLE; Ohio Integrated System Model; OISM; Reading First; Public Charter Schools; Public Preschool; Homeless; McKinney Vento; Career Technical Adult Education

Summary: ODE Competitive Grants Reader Training and Scoring Process

ODE Competitive Grants Reader Training and Scoring Process

1. Create and up-load the Request for proposal (RFP) for the competitive grants process (CCIP) according to the specifications of the electronic CCIP guidelines.

2. Draft a scoring instrument that conforms to the CCIP categories using a 6 point rubric for each item. There must be a one-to-one correspondence between the items on the scoring rubric and the material in the RFP. You should develop between 15 and 20 items that correspond to the major sections of the CCIP process.
Share the draft instrument with OSU for feedback on the items, rubric and response options. Revise as appropriate.

3. Up-load the final instrument and rubric into the CCIP electronic application so that it is attached to the RFP.

4. The CCIP process automatically tracks each potential application as it is developed, submitted and logged into the system marking each application with a time stamp.

5. Contact *Ohio State University's ODE competitive grant contact (OSU)* and alert him to the number of applications received for your program (also alert him if you have more than one type of application) and mutually arrange for a date and time for reader training, reading and scoring of applications and identify any unique features about your application(s), e.g., specific weighting of items; specific weighting of the logit score above and beyond the basic calibration.

6. Develop a calibration application that all proposal readers will read and score. The calibration application can be a real application or one that you create. Frequently, the calibration application is the highest rated application from the previous year with appropriate edits to accommodate changes in the RFP and/or dates of activities. In any event, the calibration application should have a

reasonable degree of quality so that the readers can begin to get "seated" on quality.

7. Recruit readers who have the knowledge and skill level needed to professionally read and score your applications. A rule of thumb is approximately one reader for each of the applications that you have received. The more readers you have the easier the process will go and minimize reader fatigue, keeping the number of proposals to be read by a given reader manageable. One issue that you need to be sensitive to is a possible conflict of interest by members of your reading pool. If you know or suspect this as a potential problem, discuss this with William Loadman to devise a strategy for dealing with this potential problem. Only trained readers are permitted to read and score an application.

8. Prepare for the training of readers/day by following the guidelines on the Day of Reading Material (attached).

9. The day of training will follow an **Agenda** that includes:
   - Welcome
   - Presentation of grant philosophy and RFP
   - Grant Reading Perspectives (See attached)
   - Review of Scoring Rubric items and response options
   - Review of scan sheet completion with needed designated fields (reader number, application number and item score responses). Discuss the issue of conflict of interest and have each reader sign and date the disclosure statement denying a conflict of interest on each application that they read and score
   - Discussion of providing feedback to each applicant on strengths, limitations and recommendations on Feedback form
   - Introduction, reading and scoring of calibration application
   - Discussion of calibration application strengths, limitations and recommendations getting the readers to make comments
   - Collect a subset of readers scores from the scoring rubric on the calibration application and project those to the entire group of readers (use about 12 blinded readers scores
   - Discuss the scores with the group, looking for obvious outliers, flat liners and if the grant program will engage in resolving scoring discrepancies on each item of 2 or more points, explain how this will happen and demonstrate with an example
   - Collect the scan sheets and have someone review each of these documents looking for missing information, e.g., disclosure signature, reader number, application number and a score for each item on the rubric. Have the reader add missing information as necessary

- Depending upon the number of proposals to be read, the actual "reads" may follow directly after the training session or they can be scheduled for another day and time

10. Each proposal must receive a minimum of 3 independent reads. Additional independent reads will increase the calibration on each application. Continuing to add readers beyond 5 will usually not substantially increase the precision of the calibration.

11. Depending upon the length of the proposals, the time needed to read and score a typical proposal using a typical rubric will be about 45 minutes. The CCIP process limits the amount of text in any one section of an application to 3000 characters and therefore standardizes the application length and also partially standardizes the reading and scoring procedures for the reader. Conducting discrepancy resolution among the readers of a given proposal will add 10-15 minutes to the process, but usually improves calibration and minimizes post review reaction of applicants due to high potential reader variability.

12. Resolving discrepancies among the readers of a given proposal usually functions well and readers report liking this part of the process. This process also generally minimizes potential conflicting feedback to the applicant

13. As readers complete the reading and scoring, they turn in the scan sheets and feedback form to a central location where program staff review the scan sheets for completeness and get readers to edit for missing information. This minimizes data entry errors and minimizes data cleaning procedures.

Application Calibration Procedures

14. Take the calibration scan sheets and the trios (or more) of coded scan sheets to ODE and he will get the sheets scanned and turned into a rough data file. ODE will electronically send OSU and the ODE project officer a coded data file.

15. OSU will examine the data file for obvious errors, e.g., less than three reads per application, application numbers out of field, reader numbers out of field, the correct number of applications and item scores that are out of field. If any preliminary "noise" is found in the data set, OSU will immediately contact the program manager and attempt to work with the data manager to rectify the problem, e.g., going back to the raw data to try to figure out what the missing and/or wrong information should be. If the information can be quickly corrected, the raw data file is then edited and preliminary analyses are begun. If the problem can't immediately be corrected then more aggressive measures are

initiated such as getting an additional read (by one of the trained readers) to insert into the data file. Following the corrective action, the revised data file is used to conduct the analyses. At no time during the data collection and analysis activities does the OSU representative know or have access to the name of any applicant or reader name associated with the process. OSU only has access to the reader numbers and the application numbers that have been arbitrarily applied. The program manager maintains the cross walk between the application and reader code information and the assigned application and readers' names.

16. An item response theory based program, called Facets is then applied to the cleaned data set. The program is designed to simultaneously calibrate each application, each item on the scoring rubric and each reader using a polychotomous scoring program. The simultaneously calibrations for each facet (applications, readers and items) yields an arbitrary metric called a logit that for applications yields a logit score for each application that can easily be ordered from high to low along with an accompanying standard error and a measure of reliability for being able to reliably distinguish among the applications. Further, the information generated for each application also identifies the degree of agreement /non agreement among the readers as well as any indication of flat lining associated with the scores for the application. Additional information provided is an adjusted raw score in the metric of an item rubric for each application. Using this information it is possible to identify where quality in the scored application begins to wane and to reliably set a cut score through the application of the adjusted raw score and the standard error (in the form of a root mean square error [RMSE]).

For the reader facet, a different logit score is produced for each reader. The program estimates the degree of severity/leniency for each reader and then uses that information to help calibrate the proposals read by each reader. An overall estimate of reader reliability is also obtained. Finally, estimates of the precision/error of each reader are estimated as well as any indication of potential flat lining. Any reader demonstrating a high degree of error is subject to having their scores removed and a new set of reads inserted in their place (from the set of trained readers. In more than 15 years of conducting this procedure through ODE not more than a handful of readers have been removed for cause, e.g., late discovery of a potential conflict of interest, or very high error rates in the readings. It is possible to provide feedback to the reader on their degree of leniency or severity; it is also possible to use this information to decide whether or not to invite a reader back to conduct future reads based on their performance. The Facet program is very sensitive and picks up even minor aberrations in reader, item and/or application scores.

Finally, the item facet yields yet a different logit score and the items are calibrated yielding another reliability estimate as well as another RMSE. The item calibration is also used to help adjust the scores on the other two facets. An adjusted raw score in the metric of the item is produced which is helpful in knowing which items are easier or harder to receive a higher score, which items are associated with less precision and which are prone to flat lining. These latter pieces of data can be helpful in modifying the rubric for future use and/or in deciding to potentially eliminate an item from the scoring (this almost never happens, particularly after the rubric has been previously used and calibrated).

Following the preliminary run, the output is reviewed to see if any further noise is discovered in the raw data set, e.g., an application that has less that three reads, an aberrant reader number that doesn't appear in the reader ID list, or one application that has four reads while another only has two. Frequently these problems can be identified and fixed using a little detective work, but occasionally the error cannot be fixed and an addition read or set of reads must be generated at the 11th hour. Fortunately this is the exception rather that the rule.

The Facet program can be used to give each item equal weight or it can be programmed to weight each item differentially, based on program manager direction using an underlying theoretical foundation.

16. Assuming that the final analysis converges (resulting in a unique solution based on the data set), the output is then electronically transferred to the program manager at ODE for review. OSU then arranges for a conference with the program manager to assist in interpreting the output to the data manager as well as assisting with establishing a defensible cut score for the applications (still blinded).

17. In some instances an additional weight is applied to the logit score for the applications. Sometimes a standardized need score is used to adjust the logit scores for all applications. In the most recent round of competitive grant awards from ODE, a District Efficiency Rating (DER) score was used to adjust the logit score for applicants in each funding category according to the district DER score. See the CCIP Document Library for an explanation of the DER. For the FY 09 applications this process negatively affected about 5% of the applications.

18. Following the conference with OSU the program manager sets the cut score, crosswalks between the blinded data set and their unblinded data set and provides the list of applicants to the CCIP manager for inputting the information as to which applicants will and will not receive funding.

The program grant managers follow the analyzed order of scored applications in the awarding of funds, i.e., starting at the top of the list with the highest scored (or DER adjusted) application and working down the list until the program funds are exhausted or the quality score on the application is such that there isn't sufficient quality to merit funding.

19. The applications to the CCIP are due at the end of April each year, (the last Friday in April) the reader training and application scoring occurs in mid to late May, the data scanning and cleaning occurs in early June and the calibration of applications, readers and items occur in approximately mid June. Awards are announced on the CCIP website on July 1.