Developing and Verifying the Psychometric Integrity of the Certification Examination for Imaging Informatics Professionals

Mark Raymond[1] and Paul G. Nagy[2]

The American Board of Imaging Informatics (ABII) was founded in 2005 by the Society of Imaging Informatics in Medicine (SIIM) and the American Registry of Radiologic Technologists (ARRT). ABII's mission is to enhance patient care, professionalism, and competence in imaging informatics. This is accomplished primarily through the development and administration of a certification examination. The creation of the exam has been an exercise in open community involvement with SIIM providing access to the PACS community and ARRT providing skilled psychometric support to ensure a balanced and comprehensive examination. The process to generate the exam required several years and the efforts of dozens of subject matter experts active who volunteered to submit and validate questions for the examination. This article describes the organizational and statistical processes used to generate test items, assemble test forms, set performance standards, and validate test scores.

KEY WORDS: PACS administration, statistic analysis, PACS training, education, medical

BACKGROUND

The call for clarification on the roles of a picture archiving and communication system (PACS) administrator rose to a loud chorus as PACS technology made the jump from the early adopter stage to the early majority stage at the turn of the century.[1] The technology underlying PACS underwent rapid innovation and was transformed from an engineering deployment into a true enterprise information system. Expensive and massive cathode ray medical monitors were replaced by light and thin liquid crystal displays. Dedicated PACS workstations were replaced with commercial PCs. The PACS that had originally been deployed on its own dedicated network joined the hospital grid. Web-based clients could now deploy images throughout a hospital to a much larger population of users. PACS administrators who had been trained in the technical aspects of supporting all this hardware were now exposed to new challenges in the form of hundreds to thousands of enterprise users and computers, with their own myriad of workflow, usability, and support needs. Moreover, innovations in technology reduced the cost of PACS, in turn accelerating the adoption of new PACS systems. The demand for support staff brought in new IT professionals from other industries with no clinical experience or understanding of radiology. PACS administrators found themselves performing multiple roles as application specialists, providing technical support, and managing relationships with multiple departments. The need was clear not only to identify a core set of competencies, important in the maturation of any profession, but to identify ways to assess and verify these competencies.

The Society for Imaging Informatics in Medicine (SIIM; at the time still known as the Society for Computer Applications in Radiology) responded by forming a task force in 2003 to address these issues. SIIM partnered with the American Registry of Radiologic Technologists

---

[1]From the American Registry of Radiologic Technologists, Principal Assessment Scientist National Board of Examiners, 1255 Northland Drive, St. Paul, MN 55120-1155, USA.

[2]From the Department of Diagnostic Imaging, University of Maryland School of Medicine, 22 South Greene Street, Baltimore, MD 21201, USA.

Correspondence to: Paul G. Nagy, Department of Diagnostic Imaging, University of Maryland School of Medicine, 22 South Greene Street, Baltimore, MD 21201, USA; e-mail: pnagy@umm.edu

(ARRT) in 2005 to create the examination, and American Board of Imaging Informatics (ABII) was formed. The ARRT has extensive experience in certification testing, providing psychometric expertise in the development of examinations, and administers tests in more than 15 areas related to medical imaging and radiologic technology to more than 35,000 professionals each year. Psychometrics is a field of science concerned with the theory and methods for developing, scoring, and validating educational and psychological tests. The science of psychometrics involves ensuring that exams measure what they are intended to measure (e.g., knowledge required for competence), that test scores are sufficiently reliable, and that score scales maintain a constant meaning over time.

In 2005 the newly formed SIIM Educational Committee directed the administration of a job analysis survey as an initial step toward the development of the Certified Imaging Informatics Professional (CIIP) program. The survey included 127 competency statements organized under three major domains: behavioral science, business, and technical. These domains were further subdivided into 15 subdomains or roles, with each role including six to 15 competency statements. The competency statements represented a combination of job responsibilities and knowledge, skills, and abilities (KSAs). Each competency was believed to be requisite for successful job performance; the purpose of the survey was to verify this assumption and to prioritize competencies. Table 1 summarizes the structure and content of the questionnaire. The

data supported the importance of most competencies, with mean ratings on the 7-point Likert scale ranging from 4.2 to 6.4. The results of the survey were published in 2005.[2] The 127 identified competencies served as the basis for the test specifications. Test specifications describe the content to be covered by an examination, the emphasis allocated to each topic, and other important features of the test. Not only are test specifications used by ABII to ensure that different forms of an exam are comparable over time, but they also serve as a useful guide for examinees, those who provide training, and employers who wish to know the specific competencies covered by the test.

The SIIM Certification Committee and ARRT staff worked on the test specifications or test content outline (TCO) over a period of several months in 2006. An early draft of the TCO accompanied by a questionnaire was posted on the SIIM Web site and distributed at the SIIM annual meeting in April 2006. The questionnaire was completed by 226 individuals. Comments were generally positive, but it was evident more work specificity was needed. The Certification Committee considered alternative frameworks and ultimately settled on an outline consisting of 10 major domains, with three to six specific job responsibilities listed under each major domain. Within each specific responsibility, the outline further identified the KSAs required to effectively carry out those responsibilities. The 10 major domains and their associated weights appear in Table 2. In October 2006, the Certification

Table 1. Initial Job Analysis Questionnaire: Structure and Sample Competencies

Domain	Role	Sample competency
	Training	Developing user training programs
	Workflow engineering	Workflow analysis
	Reading environment	RIS–PACS dictation integration
Behavioral	Customer relations management	Overcoming psychological barriers
	PACS readiness	Understanding the CIO perspective
	Strategic visions	Building strategic and operational committees
	Economics of PACS	Total cost of ownership
	Vendor selection	Vendor support
	Project management	Performance milestones
Business	Sustaining PACS	Recruiting PACS professionals
	Technology overview	Workstations and displays
	Systems management	Recoverability policies
	Troubleshooting	Network administration
	Modalities	Integration with PACS via DICOM/IHE
Technical	Security	Understanding HIPAA security and auditing

Table 2. Major Performance Domains Comprising the TCO

Performance domain	Portion of exam (%)
Procurement	5
Project management	5
Operations	10
Communications	10
Training and education	5
Image management	20
Information technology	15
Systems management	10
Clinical engineering	10
Medical informatics	10

One hundred fifty questions (130 scored and 20 pilot)

Committee approved the TCO and also determined that the exam would include 150 items (130 scored plus 20 unscored pilot items). A complete copy of the TCO is available at abii.org.[3]

Examination Development

Launching and maintaining a certification examination requires the availability of hundreds of high-quality test items addressing each area of TCO. Representatives from the Certification Committee and general SIIM membership attended the first item writing workshop held in July 2006 at the ARRT offices in St. Paul, MN. The goals of the workshop were to obtain input on the TCO, which was still in draft form, to learn basic principles of test item writing, and to begin building up a pool of test items. A second workshop held, in February 2007, was attended by six individuals from the SIIM membership. Approximately 400 questions were submitted during and subsequent to these first two workshops. Newly written test items are submitted using secure Web-based software written specifically for item writing and review. Once submitted, items undergo two levels of review. The first level is completed by a review panel including item writers and other volunteer reviewers. Items are rated for relevance, technical accuracy, and overall quality. The second level of review is completed by members of the Examination Committee, who review the items in light of comments and ratings provided by initial reviewers. Items at the second level of review are either accepted as is, accepted with revision, or rejected. Once accepted, items migrate into a database called an item bank. The item bank contains hundreds of fields of informa-

tion organized into various tables that facilitate each item's use on multiple test forms over time. In addition to the question itself, the item bank stores information such as the item's author, reference, date of acceptance, edit history, classification under the TCO, accompanying graphics, usage history, numerous statistical indices, and other data. To ensure an ample supply of test items for new forms of the exam and to make certain that future exams keep pace with advances in technology and changes in job responsibilities, new item writers are appointed and additional workshops are periodically held. The Certification Committee convened two meetings early in 2007 to begin the development of the test form to be used for the pilot study. The first meeting was held in January and was devoted primarily to reviewing, revising, and classifying newly submitted test items. At the conclusion of this meeting a tally of the item bank was completed, and item writers were provided with feedback regarding content areas in need of test questions. A second 3-day meeting was held in April 2007, during which the Committee continued its review and revision of new questions. It was also necessary for the Committee to write questions to obtain specified coverage of all content areas. The pilot study exam form was edited and finalized by psychometric staff at ARRT. It was administered to 100 qualified participants as a paper-and-pencil test on June 9, 2007, at the SIIM annual meeting in Providence, RI. Examinees were invited to comment on individual questions for relevance and accuracy and to complete a short survey regarding the exam as a whole.

Statistical Analysis and Standard Setting

Immediately after the pilot test, responses to each question were subjected to an item analysis to verify the accuracy of the scoring key and to statistically validate each question. Although a great deal of effort goes into the writing of each question, the real test is in determining how appropriate, clearly written, and effective that question was. An additional purpose of the item analysis was to help determine which 130 items (of the 150) would be used for scoring purposes. Examinee comments were also evaluated as part of this decision-making process. A statistical item analysis evaluates responses to each question as a
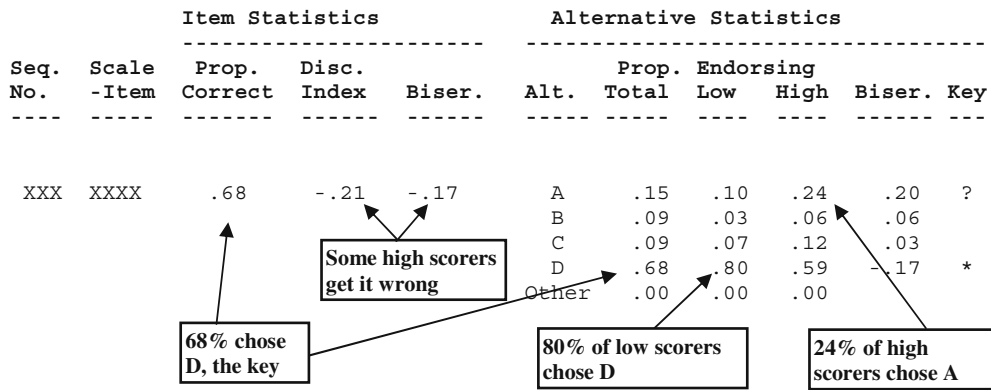
```
                Item Statistics                       Alternative Statistics
                ----------------------                ------------------------------------
   Seq.  Scale  Prop.   Disc.                                 Prop. Endorsing
   No.   -Item  Correct Index   Biser.   Alt.  Total  Low   High  Biser. Key
   ----  -----  ------- ------  ------   ----- -----  ----  ----  ------ ---


   XXX   XXXX    .68     -.21    -.17      A    .15    .10   .24    .20    ?
                                           B    .09    .03   .06    .06
                                           C    .09    .07   .12    .03
                                           D    .68    .80   .59   -.17    *
                                         Other  .00    .00   .00
```

Some high scorers get it wrong

68% chose D, the key

80% of low scorers chose D

24% of high scorers chose A

Fig 1. Sample output from statistical item analysis.

function of each examinee's total test score. In general, individuals with high overall scores are expected to have a higher probability than examinees with low scores of answering a specific item correctly. When a significant portion of high-scoring test takers respond to an item incorrectly, the item is scrutinized for accuracy and clarity of wording. To illustrate the utility of an item analysis, Fig. 1 represents the output for a single test item with questionable statistics. Although most individuals correctly answered this sample item, many high scorers answered it incorrectly, as indicated by the negative discrimination index, negative biserial correlation,[4] and relatively high proportion of high-scoring examinees (24%) choosing option A. Statistics such as this do not always mean that the item is flawed, but they work well for screening purposes. In the case of the item in Fig. 1, the item was worded such that option A also appeared correct. Flawed items can either be replaced with a pilot item, given multiple correct answers, or simply removed from scoring. The Certification Committee might later revise such items or discard them from the item bank.

Other examination statistics are also evaluated to verify that the test is functioning as intended, including the mean, standard deviation, range, frequency distribution, and other graphical displays. Three very important values are an index of reliability (such as coefficient alpha), the standard error of measurement, and an index of decision consistency. All provide information about the consistency or dependability of test scores. Passing scores for most certification examinations are established using criterion-referenced standard-setting procedures. Examinees pass or fail based

entirely on whether their level of proficiency as measured by the examination meets or exceeds some absolute criterion, or cut-off score. It is—but unlikely—that everyone could pass or that everyone could fail. This stands in sharp contrast to a norm-referenced test, where some predetermined percentage of examinees will pass or fail regardless of their proficiency.

A criterion-referenced standard-setting procedure known as the Angoff method was used to set the passing score for the CIIP examination.[5] The standard setting study was conducted about 2 weeks after the pilot exam was administered. The eight participants included members of the Certification Committee, as well as other representatives from the imaging informatics community. The Angoff method requires that participants inspect each item on the exam and estimate the proportion of minimally qualified test takers who would get the item correct. This estimated proportion is called an item rating. The sum of the item ratings for a participant is an expected score for a minimally qualified examinee. The objective of the Angoff method is to approximate the empirical outcome that would be obtained if it were possible to actually give the exam to a group of individuals who had been determined, a priori, to be mini-

Table 3. Exam Performance, 2008

Exam date	No.	Min	Max	Mean	SD	% pass
June 2007	103[a]	59	97	85.7	6.4	96.1
Sept 2007	96[a]	56	97	83.1	7.6	87.5
March 2008	103	64	98	83.0	7.7	84.5
Sept 2008	120	61	94	80.8	7.0	80.8

[a]Data exclude members of the Certification Committee

mally qualified. By asking multiple participants to make their ratings independently, it is reasonable to average their scores to derive a passing standard, which represents a consensus of agreement on the expected performance of minimally qualified candidates. Scores were converted via linear transformation such that the scaled passing score would correspond to 75 and the maximum scaled score would be 99. Of the 103 examinees who took the pilot exam, 99 passed. The high pass rate was attributed to the fact that a majority of individuals who signed up for the pilot study were highly motivated and very proficient early adopters.

Subsequent Examination Forms

Four additional forms of the CIIP examination have been administered in the 2 years since the pilot study. Throughout this time frame, item writers continued to produce new items, and the Certification (Examination Committee) continued meeting to review items and assemble new test forms. To guarantee that the new test forms are comparable in terms of test content, all forms have been constructed in strict accordance with the TCO. To ensure that scores on different exam forms are comparable, statistical equating is used.[6] Equating requires that a new exam form have about 20–25% of its items in common with some previous form; then statistical models can be used to detect and correct for differences in test form difficulty. So, for example, if test B is harder than test A, then test B would have a lower passing point on the raw score scale. However, scores from both forms would be placed on the same scale for reporting purposes—a scale where the passing point is set at 75.

Table 3 summarizes results for the first four test administrations. The slight decrease in mean scores and pass rates and increase in variability suggest that the initial pilot study group probably was not representative of the entire population of imaging informatics personnel (i.e., they were more proficient). However, although the pass rate has dropped, the current pass rates are reasonable and indicate that examinees are generally quite proficient. For example, pass rates for other certification programs in medical imaging typically range from about 80% to 90%.[7,8]

CONCLUSION

The journey to create an examination to evaluate imaging informatics professionals has taken several years and hundreds of volunteer hours. The process is slow but deliberate, involving practicing informaticists in scrutinizing closely the results of each test to ensure we are providing a useful instrument for professionals to improve their careers. ABII frequently invites diplomates of the exam to submit new items for future tests and become involved in exam assembly working groups. The industry of imaging informatics in medicine is still undergoing powerful transformations as information technology has escalating effects on the delivery of care. By working closely with practicing informaticists, ABII has built an organizational engine that can evolve and adapt the examination to the changing needs of our industry.

ACKNOWLEDGMENT

REFERENCES

1. Reiner B, Siegel E: Psychological factors affecting the adoption of PACS. Appl Radiol 31:4, 2003

2. Nagy P, Bowers G, Reiner B, Siegel E: Defining the PACS profession: an initial survey of skills, training, and capabilities for PACS. J Digit Imaging 18(4):251–259, 2005

3. The American Board of Imaging Informatics Test Content Outline. Available at http://www.abii.org/ABII/faces/abii_test_content_outline.jsp Accessed 1 May, 2009

4. Lord FM, Novick MR: Statistical Theories of Mental Test Scores. Reading, MA, 1968

5. Cizek GJ: Setting Performance Standards. Mahwah, NJ: Lawrence Erlbaum Assoc, 2001

6. Kolen MJ, Brennan RL: Test Equating, Scaling, and Linking: Methods and Practices. New York, NY: Springer-Verlag, 2004

7. American Registry of Radiologic Technolgists. Annual Report of Examinations. St. Paul, MN: ARRT, 2008

8. American Board of Radiology. Available at http://abr.org Accessed 1 May 2009