e-profiling: An Alternative to Traditional Assessment Methods

Canadian Environmental Certification Approvals Board
1999-08-04

Table of Contents

1. Background

A new assessment tool called e-profiling, which combines the best practices of Canadian credentialing bodies with the advantages of Internet technology, is discussed in terms of its strengths and weaknesses. The strengths and weaknesses of alternative assessment methods are also outlined.

In Canada, credentialing bodies use three assessment methods: direct observation, simulation, and written/oral testing. Of the three, no one assessment method is better than other methods. Selecting among them involves a tradeoff between the different strengths and weaknesses of each. Whereas the weaknesses of written/oral testing and simulation are primarily due to factors other than economical or logistic, one of the main drawbacks of direct observation is due to economic or logistical factors. By incorporating the improved efficiency and economics of Internet technology into a direct observation assessment model, previous weaknesses are improved upon. While new weaknesses arise with e-profiling, policy decisions, quality assurance and quality control measures minimize their impact. As with the current assessment methods, the combination of Internet-technology with direct observation has strengths and weaknesses. As such, e-profiling is an alternative to the existing methods; where, selecting among the alternative assessment methods involves selecting among the respective strengths and weaknesses of each.

Before expanding on the above points, a brief outline of the factors motivating the development of an alternative, equally rigorous assessment method is provided.

2. Why develop an alternative assessment method to existing methods?

The improved efficiency and economics of e-profiling not only improves upon a main weakness of direct observation, e-profiling improves the viability of an industry-initiated, practitioner-lead, voluntary certification.

There is a documented need for on-going certification in the environmental sector. Within the dynamic, knowledge-based environmental sector, a company's competitive advantage comes from the skills and knowledge of its workers. There is widespread recognition among sector practitioners, companies and clients that a reputable voluntary credential with on-going professional development is a desirable prerequisite for commissioning environmental work.[1] Moreover, the changing technological and legislative base of the sector requires that the certification have a strong on-going professional development component. An industry-initiated, practitioner-lead, voluntary certification, called the Canadian Certified Environmental Practitioner (CCEP), has been developed to fulfil these needs, under the auspices of the Canadian Environmental Certification Approvals Board (CECAB).

While there is a demand for certification, there is a limit as to what candidates will pay. Given the voluntary nature of the certification, if the cost of certification adversely affects the capability of the largest component of the sector (small and medium sized firms) to compete, in what is a cost-competitive environment, the cost of voluntary practitioner certification will be difficult to justify.[2] Hence, the demand for voluntary certifications is price sensitive. Unlike licensure, voluntary certifications afford a competitive advantage and not a monopolistic right to practice.

Due to the ceiling price, candidates pay only a fraction of the actual cost of certification. This fact is prohibitive to the success of new certification bodies. While existing bodies defray high operational costs through volunteers, this option is not open to new certification bodies that have yet to acquire the necessary membership base. The losses incurred during the time taken to obtain a large enough membership base, can be detrimental to success. If the benefits of a reputable voluntary certification program (with ongoing professional development) are to be realized by practitioners, companies, and clients in the environmental sector, an assessment method, as rigorous as existing methods, with improved economics and efficiencies is required. Mindful of such considerations, CECAB designed the e-profiling system.

A discussion of e-profiling, as an alternative, equally rigorous certification method to existing methods, follows.

[1] For a more complete discussion and study concerning the demand for certification see: Canadian Council for Human Resources in the Environment Industry (CCHREI) (1999). *A Profile of the Canadian Environment Industry and Its Human Resources*.

[2] According to CCHREI (1999), small firms with 2 to10 employees account for 50% of employment in environmental industry. Firms with between 11 to 50 employees account for 31% and sole proprietorships account for 2%. In total, 83% of all employment in the environmental industry is accounted for by small and medium enterprises (SME's).

3. Three Methods of Assessment for Certification

Certification bodies evaluate competence using various assessment methods, including direct observation, simulation and (written or oral) examination. Direct observation typically involves trained observers who review the candidate's performance with real clients, using rating scales or checklists. During a simulation, the second assessment method, the candidate is provided with a description of the client, the client's problem and the circumstances surrounding the encounter. Questions concerning potential actions are then presented to the candidate; each choice elicits further choices. The final mode of assessment, written or oral examination, is familiar. Oral or written examination involves an evaluation of the candidate's capability by having the candidate respond, either in writing or verbally, to written questions or those asked by a panel of examiners.

3.1. Selecting a Method Involves a Tradeoff of Strengths and Weaknesses

The use of assessment results to render valid credentialing decisions requires three inferences; whereby, a serious flaw in any of the three inferences can invalidate the use of the results (Kane, 1992b). The first inference requires that the data collected be a representative sample of the candidate's general level of performance. The second inference requires that the content of the assessment be relevant to the actual practice, in which the credential is sought. Finally, a clear credible basis for differentiating good performance from bad must be used during the assessment. Each assessment method has strengths and weaknesses with the inferences; selecting among them involves a tradeoff between their respective strengths and weaknesses (Kane, 1992b).

3.2 Direct Observation

With direct observation, subjectivity may be involved in assigning scores. Real-world situations are often not 'textbook'. Not only are 'best' solutions not always immediately apparent, but 'experts' often disagree about the 'best' solution(s) to 'non-textbook' situations. The conditions permit variations in interpretation and hence variations in scoring (Hoffman, 1977; Hubbard, 1971).

In addition, a candidate's variability in performance tends to be large and the number of observations tends to be small. As a result, generalizing from a small, possibly unrepresentative, sample is a potential threat to direct observation (Swanson, 1990). The gathering of samples tends to be small in number, over a limited time period and not very broad in terms of content for economic and logistical reasons. Kane (1992b) notes that "observing performance in actual practice is sufficiently inconvenient and expensive that the samples of performance are very small. Data collection typically occurs over a limited period of time, during which the examinee works with a limited number of clients in a specific context."

On the positive side, direct observation provides an assessment of actual performance, and therefore the relevance of what is assessed to actual practice is high. As one might expect, the relevance of the content assessed to actual practice is the strongest inference for direct observation.

3.3 Simulation

With simulation, the aim is to make each encounter as 'real' as possible, while controlling for variables, presenting measurement difficulties for the direct observation method (Kane, 1992b). A trade-off exists between achieving a simulation that is standardized to promote reliability in scoring, and yet realistic. As encounters become more standardized, they become less realistic and tend not to reflect actual practice.

The more realistic the simulation, the less likely expert raters are to agree on scoring (Swanson, 1990). According to Swanson (1990) "it is difficult to develop scoring strategies that appropriately reward alternative strategies that are equivalent in quality. It is also difficult to ensure that similar strategies differing in quality receive appropriate score." Additionally, testing tends to be for thoroughness and not efficiency (Kane, 1992b).

As with direct observations, simulations can suffer from both a small sample size and variability in performance. However, because the set of simulations can be designed to assess a candidate over a broader range of situations, results from a simulation tend to be more reliable for generalizing.

Since simulations do not, by definition, involve real clients there is the possibility that performance in a simulation may not reflect actual performance. Feightner (1985), Goran, *et al.* (1973) and Page and Fielding (1980) all reported a lack of correlation between performance recorded with a simulation and actual performance. As noted above, testing is usually for thoroughness and not efficiency. Efficiency may be the norm in actual practice.

3.4 Oral or Written Testing

Test development often requires a panel of experts to agree upon the questions to be included in the exam. As a result, there is a tendency for tests to contain factual or knowledge-based questions. An explicit effort must be made to include judgement questions in exams (Swanson, 1990). Written or oral tests provide, at best, a direct measure of enabling skills and knowledge, but an indirect indication of performance in actual situations (Kane, 1994, 1992a,1992b, 1982; Ongley, 1970). Hence, the inference from test scores to actual performance is a substantial weakness for testing and is a threat to the validity of interpreting testing results when rendering credentialing decisions.

On the positive side, because a large quantity of questions and a wide range of content can be efficiently covered, generalization from the sample is considered the strongest inference for testing.

Awareness of the strengths and weaknesses, inherent in the alternate assessment methods, not only assists in selecting among the methods, but knowledge of the weaknesses is vital for developing a new assessment method. The areas requiring extra attention, given the characteristics and general limitations of a particular method, can be identified and given the consideration they deserve.

4 Why e-profiling Uses a Direct Observation Mode of Assessment

Whereas the weaknesses with written/oral testing and simulation modes of assessment are primarily due to factors other than economical or logistic, a major weakness of direct observation is due to economic and logistical limitations. As noted, the high cost and inconvenience of directly observing a candidate's performance with actual clients are the main factors limiting the number of observations, the time period over which the observations are made and the breadth of content assessed; hence, they limit how representative the sample is. Generalizing from a small, possibly unrepresentative, sample is a threat to direct observation and, therefore, an area requiring consideration. A more efficient and economical means of performing direct assessments would enable a larger more representative sample, strengthening the generalization inference.

Technological advances improve the efficiency and costs of doing business. The Internet is no exception; it has brought both efficiencies and improved economics to commerce. Capitalizing on the improved efficiency and economics, the Canadian Environmental Certification Approvals Board (CECAB) has developed an on-line certification system. The system, called e-profiling, integrates the best practices of existing certification bodies, with the advantages of the Internet. The result is a direct assessment tool that minimizes some of the weaknesses previously due to economic and logistical constraints. A brief digression, outlining the e-profiling process, will be useful prior to investigating the result of combining Internet technology with a direct observation assessment method.

4.1 What is e-profiling?

While e-profiling uses several algorithms to compare a candidate's responses against both an occupational standard, and later peer evaluator responses, the assessment process requires both on-line and substantial off-line activities. It is estimated that a typical assessment will require a candidate to spend approximately an hour and a half (1.5) on-line and seven and a half (7.5) hours off-line.

Consistent with the best practices of credentialing bodies, e-profiling has two stages: the candidate's declaration of competencies followed by a validation stage. A candidate begins the process by downloading general background material, a list of the requirements for the credential, as well as an overview of the e-profiling process. Once a decision to proceed is made, the candidate goes on-line and registers, providing personal information, an employment history, academic background and membership in professional or industry associations. In addition, the candidate selects the area (or areas) in which certification is sought. After various statements of confidentiality and declarations, attesting to the completeness and accuracy of the information, are signed and received by CECAB, the candidate is e-mailed material to complete a self-assessment. In addition, information and support material, to assist with identifying and securing qualified evaluators, is supplied to the candidate.

When completing the self-assessment, the candidate responds to a set of questions for each competency statement. Depending on the area being assessed the number of competency statements range from 41 to 85. The questions answered concern: the amount of experience by

the candidate, the last time that the competency was used, and finally the degree to which the competency is performed successfully without assistance.

Candidates are likely to be accurate concerning the amount of experience they have and the last time that a competency was used; however, it has been empirically shown that individuals typically underestimate their level of ability when it is known that statements will be verified. Candidates are made aware that self-declared statements will be verified.

The candidate's application proceeds to the second stage - the validation stage - when the self-declared competencies meet or surpass certification requirements. As noted above, certification requirements are based on national occupational standards (NOS) developed and kept current by the Canadian Council for Human Resources in the Environment Industry (CCHREI).

To determine if an application proceeds to the validation stage, an algorithm is run and a gap analysis is performed. On the basis of a candidate's input, the algorithm establishes if the candidate meets the certification's requirement; the gap analysis identifies both the candidate's strengths and areas of improvement. If a candidate does not have the required competencies, to proceed to the validation stage, the candidate is informed which competencies need to be acquired and asked to reaffirm that the initial input was accurate. If the candidate chooses, the gap analysis can be used in conjunction with a CECAB on-line professional development service. The candidate can construct a personalized professional development program, monitoring and recording progress *en route* to certification.

It is worth noting that the candidate’s self-declared responses could be validated using any of the three assessment methods discussed above. For example, the self-declared responses could be used to construct a written or oral exam specific to the candidate. Alternatively, the self-declared responses could be used to construct a set of simulations for the candidate to be tested against. However, since Internet technology offers efficiencies that improves upon a general weakness for direct observation methods (whereas the weakness of written/oral examination or simulations are due to factors other than economic or logistic), a direct observation mode of assessment was selected.

At the verification stage of a candidate's competencies, e-profiling uses a consistent methodology in which independent, third-party peer evaluators perform the validation. Steps are taken to provide for the anonymity of the peer evaluators and the confidentiality of all information supplied to CECAB. Peer evaluators (all of whom fit the criteria and have a history within the past five years with the candidate) answer questions concerning the candidate's demonstrated environmental capability. An algorithm is run and the responses of the independent peer evaluators are compared against the candidate's responses. If the peer evaluators do not confirm the candidate's self-assessed responses, the candidate has the right to appeal. CECAB provides an on-line mechanism for appeals in addition to instructions for an alternative form of appeal. A practitioner that meets or surpasses the requirements will be issued a certificate and referred to as a Canadian Certified Environmental Practitioner and be denoted by CCEP.

5. e-profiling: A Review of Its Strengths and Weaknesses

As noted, e-profiling employs a peer evaluation mode for direct assessment. According to the literature, peer evaluation is a valid assessment method often used in high-stake assessment situations such as those found in medical schools (Norton, S., 1992). Burnett, W. and Cavaye, G. (1980) report that peer evaluation, used in medical residency programs, is a reliable form of assessment. Over twenty years of findings are fairly consistent across peer evaluation studies. For example, Morton, J. B. and Macbeth, W. A. A. G. (1977) reported that when compared to self-assessments and instructor assessments, peer assessments were lower than medical staff and that self-assessments were lower than peer assessment. In 1998, Shore, L *et al.,* reported that the scores among the three were fairly close.

In addition to medical schools, peer evaluation is used for hiring practices, performance assessment and employee development in industry (Bettenhausen, K.L., and Fedor, D.B., 1999), as well as by certification bodies. In the case of the latter, practitioners previously certified by the process, evaluate new candidates in at least one of two ways. First, a practitioner may **directly review** the candidate's performance. Second, a practitioner who works with the candidate may **sign-off** as to the candidate's capability based on a history of working with the candidate.

E-profiling differs in at least two important aspects with traditional methods for direct observation. First, the national occupational standard plus some of the candidate's self-assessed responses provide the content for the assessment. As a result, each assessment is tailored to the candidate while maintaining that a core set of competencies, required of all practitioners, is assessed. [The flexibility of this approach is required in an occupational sector where diverse skill-sets are the norm and not the exception.]

The second manner in which e-profiling differs from other direct observation methods concerns the fact that peer evaluators are used. Discussion of the second difference occurs below.

The differences in the e-profiling process improves upon some of the weaknesses common among direct observation assessment methods. It was noted above that the difficulty of generalizing from a potentially unrepresentative sample is a general weakness with direct observation methods of assessments. However, unlike other direct observation assessments, a more representative sample is obtained using e-profiling. As a result, the inference is strengthened.

Logistics and economics were identified as factors limiting the number of observations in the sample, the breadth of content observed for a sample, and the time span over which observations are made. However, through the use of Internet technology and peer evaluation, e-profiling reduces the role that these factors have on the sample size by allowing a greater number of observations and a greater time period over which samples are gathered. The breadth of experience that five peer evaluators, over a five-year period, have with a candidate's work is greater than the few observations made by one or two observers over a limited time period. Various policies concerning evaluation, along with quality assurance (QA) and quality control (QC) measures effectively control the quality of the evaluators, the number of evaluators and the time period that evaluators are familiar with the candidate's work.

In addition, the number and type of competencies selected for peer evaluation control both the content and breadth of the assessment. As noted above, selecting an assessment's content is a function of both the occupational standard and the applicant's self-declared strengths. As a result of this and e-profiling's peer evaluation model, the breadth of the assessments is controlled by policy decisions and not the logistic and economic factors, which usually limits the content assessed to the work performed during the observer's visit.

A second weakness, identified by Kane (1992b) as a common weakness for direct observation, concerns the potential for subjectivity. When peer evaluators are used the potential for subjectivity is also present. Factors contributing to peer evaluator subjectivity include duress, confidentiality and the fact that peers are used for assessment. How subjectivity is managed within a peer evaluator process will be addressed after a discussion of duress and confidentiality.

Duress is addressed by creating conditions that provide for the anonymity of peer evaluators. While five evaluator's are supplied by the candidate, the credentialing body retains the option to select how many and which evaluators perform the assessment as well as the right to request alternative evaluators. Activity or lack of activity between a peer evaluator and the credentialing body is never released. Moreover, the candidate signs a waiver agreeing not to seek directly or indirectly any information concerning the activities of evaluators, as well the candidate agrees to hold harmless the peer evaluators, and all associates of the credentialing body and the body itself for actions (or the lack of) resulting from the certification process. In return, the peer evaluators also agree not to reveal material facts, including whether or not they have performed an assessment. CECAB likewise agrees not to release information. Together, these conditions contribute to an environment of anonymity and confidentiality for all concerned.

The subjectivity that arises during an assessment is an issue for all assessment methods. Subjectivity cannot be eliminated; at most, it can be controlled (Kane, 1992b). How is subjectivity controlled for in e-profiling? As with other direct observation methods, e-profiling uses a form of rating scale to control subjectivity. At least five factors assist in minimizing subjectivity. Together these factors contribute to agreement among the scoring of independent raters. The agreement among raters is contrary to what was reported by Hoffman (1977) and Hubbard (1971). Each study reported a lack of agreement among raters. Kane (1992b) concludes that subjectivity, as evidenced by the lack of agreement, is a common problem for direct observation assessments. As noted, e-profiling obtains agreement among independent raters.

The factors are as follows: First, the national occupational standards (NOS), which form the content of assessment, are grass roots in origin. That is, the NOS originated from experienced, knowledgeable environmental practitioners. Second, the NOS is presented in terminology familiar to environmental practitioners. As a result, there is less likelihood that practitioners will be unclear about the competency in question. Third, active, experienced and knowledgeable practitioners perform the assessment. Together, the first three factors reduce subjectivity by having experienced and knowledgeable practitioners, who know what it is to perform a task at the level required by industry, assess candidates against competency statements, written in

terminology used by environmental practitioners. After all, the national occupational standard originated with and is maintained by input from experienced, knowledgeable practitioners.

Fourth, the peer evaluators have direct knowledge of the candidate's work. The fourth factor fulfills an accepted criterion required for valid peer evaluation. According to Norton (1992), peer evaluation requires at least these two conditions to be valid: (1) peer's must have time to become familiar with each other's work, (2) objectively measured criteria must be used. The familiarity of the candidate's work comes from direct knowledge of the candidate's work.

The Norton's second criterion is addressed by the e-profiling's fifth factor to control evaluator subjectivity: An objective behavioural-based criterion is used in the assessment and not a subjective rating scale. Each evaluator is asked about the degree of independence the candidate exhibits when successfully completing the task. For example, the evaluator will assess the candidate as either having no experience with the competency, learning the competency, requiring occasional guidance for its successful execution, works independently or is a master in executing the competency. By asking such questions, e-profiling seeks input concerning observed behaviours and not an overtly subjective assessment of how skilled the individual is on a scale of one to five.

All the subjectivity in direct observation or any assessment method, for that matter, cannot be eliminated. The acceptability of any method rests on the degree to which subjectivity is controlled. Since e-profiling requires and obtains agreement among raters, contrary to the findings reported by Hoffman and Hubbard, there is reason to hold that e-profiling adequately controls, but does not eliminate subjectivity in the assessment.

Many of the previous weaknesses, common to direct observation methods, are minimized with e-profiling. The e-profiling method does not overcome all the weaknesses of direct observation and new issues arise with the use of a peer evaluator model, however many of the issues are controllable with proper policy and QA and QC. In all, e-profiling ranks among the current certification methods. As with these methods, each has strengths and weaknesses, but each allows for valid interpretation of results to render credentialing decisions.

Bibliography

Bangert, A. William (1996). Peer Assessment: An Instructional Strategy for Effective Implementing Performance-based Assessments. *Dissertation Abstracts International Section A: Humanities & Social Sciences.*

Bettenhausen, K. L. and Fedor, D. (1997). Peer and Upward Appraisals: A Comparison of Their Benefits and Problems. *Group And Organization Management*, 22(2): 236-263.

Bloch, B., *et al.* (1994). *Working Towards Best Practices in Assessment. A Case Study Approach to Some Issues Concerning Competency-Based Assessment in the Vocational Education and Training Sector.* South Australia: National Center for Vocational Research.

Bott, P. (1996). *Testing and Assessment in Occupational Technical Education*. Old Tappan: Prentice Hall.

Boud, David and Falchikov, Nancy (1989). Quantitative Studies of Student Self-Assessment in Higher Education: A Critical analysis of Findings. *Higher Education*, 18: 529-549.

Burnett, W. and Cavaye, G. (1980). Peer Assessment by Fifth Year Students of Surgery. *Assessment in High Education,* 5: 273-278.

Canadian Council for Human Resources in the Environment Industry (CCHREI) (1999). *A Profile of the Canadian Environment Industry and Its Human Resources*. Calgary, Canada: CCHREI.

Cole, J. R. and Cole, S. (1981). *Peer Review in the National Science Foundation: Phase II*. Washington: National Academy Press.

Conway, R., *et al*. (1993). Peer Assessment of an Individual's Contribution to a Group Project. *Assessment and Evaluation in Higher Education*, 18: 45-56.

Conway M. and Lambkin, Angela (1994). *Systems and Procedures of Certification of Qualifications in Ireland.* Berlin, Germany: European Center for the Development of Vocational Training.

Davis, J. and Inamdar, S. (1988). Use of Peer Ratings in a Pediatric Residency. *Journal of Medical Education*, 63: 647-649.

Davis, J. (1986). Interrater Agreement and Predictive Validity of Faculty Ratings of Pediatric Residents. *Journal of Medical Education*, 61: 901-905.

Davis, J. Kent and Rand, D. C. (1980). Self-Grading versus Instructor Grading. *Journal of Educational Research*, 73: 207-211.

Doyle, J. R. and Green, R. H. (1994). Self and Peer Appraisal in Higher Education. *Higher Education*, 28: 241-264.

Edwards, A., *et al.* (eds.) (1995). Assessing Competence in Higher Education. Staff and Educational Development Series. *ERIC Document Reproduction Services* No. ED387048.

Elliott, Stephen, N. (1995). Creating Meaningful Performance Assessments. *ERIC Digest.* Eric Document Reproduction Service No. ED381985.

Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education and Training International*, 32: 175-187.

Falchikov, N. (1988). Self and Peer Assessment of a Group Project Designed to Promote the Skills of Capability. *Programmed Learning and Educational Technology*, 25: 327-339.

Falchikov, N. (1986). Product Comparison and Process Benefits of Collaborative Peer Group and Self Assessments. *Assessment and Evaluation in Higher Education*, 11: 146-166.

Falchikov, N. and Magin, D. (1997). Detecting Gender Bias in Peer Marking of Students' Group Process Work. *Assessment and Evaluation in Higher Education*, 22: 385-396.

Feightner, J. W. (1985). Patient Management Problems. V. R. Neufeld and G. B. Norman (eds.) *Assessing Clinical Competence*. New York: Springer Publishing Company. pp. 183-200.

Feurer, I. D., Becker, G. J., *et al.* (1994). Evaluating Peer Reviews: Pilot Testing of a Grading Instrument. *Journal of the American Medical Association*, 272(2):98-100.

Fox, S., Ben-Nahum, Z. and Yinon, Y. (1989). Perceived Similarity and Accuracy of Peer Ratings. *Journal of Applied Psychology,* 74(5): 781-786.

Goldman, R. L. and Ciesco, E. (1996). Improving Peer Review: Alternatives to Unstructured Judgments by a Single Reviewer. The Joint Commission Journal on Quality Improvement, 22 (11).

Goldman, R. L. (1994). The Reliability of Peer Assessments: A Meta-Analysis. *Evaluation and the Health Professions*, 17: 3-21.

Goran, M. J., Williamson, J. W., and Gonnella, J. S. (1973). The Validity of Patient Management Problems. *Journal of Medical Education*, 48: 171-177.

Gordon, M. J. (1991). A Review of the Validity and Accuracy of Self-Assessments in Health Professions Training. *Academic Medicine,* 66: 762-769.

Griffin, P. (1995). Competency Assessment: Avoiding the Pitfalls of the Past. *Australian and New Zealand Journal of Vocational Education Research*, 3: 34-59.

Harrington, T. F. (1995). Assessment of Abilities. *ERIC Digest*. ERIC Document Reproduction Service No. ED389960.

Hoffman, P. (1977). Continued Competence Assurance: Some Research and Measurement Considerations. *Proceedings of a National Conference for evaluating competence in the Health Professions*. New York: Professional Examination Service.

Hubbard, J. P. (1971). *Measuring Medical Education: The Tests and Procedures of The National Board of Medical Examiners*. Philadelphia: Lea & Febiger.

Hughes, I. E. and Large, B. J. (1993). Staff and Peer-Group Assessment of Oral Communication Skills. *Studies in Higher Education*, 18: 379-385.

Kane, Michael (1997). Model-Based Practice Analysis and Test Specifications. *Applied Measurement in Education*, 10: 5-18.

Kane, Michael (1994). Validating the Performance Standards Associated with Passing Scores. *Review of Educational Research*, 64: 425-461.

Kane, Michael (1993). The Validity of Performance Standards. ERIC No. ED365689.

Kane, Michael T. (1994). Validating Interpretative Arguments for Licensure and Certification Examinations. *Evaluation and the Health Professions*, 17: 133-159.

Kane, Michael T. (1992a). The Assessment of Professional Competence. *Evaluation and the Health Professions*, 15: 163-182.

Kane, Michael T. (1992b). The Validity of Assessments of Professional Competence. ERIC No. ED343958

Kane, Michael T. (1992c). An Argument-based Approach to Validity. *Psychological Bulletin*, 112: 527-535.

Kane, Michael T. (1990). Generalizing Criterion-Related Validity Evidence for Certification Requirements across Situations and Specialty Areas. ERIC No. ED337463

Kane, Michael T. (1989). Combining Data on Criticality and Frequency in Development Test Plans for Licensure and Certification Examinations. *Journal of Educational Measurement*, 26(1): 17-27.

Kane, Michael T. (1982). The Validity of Licensure Examinations. *American Psychologist*, 37: 911-918.

Kutner, Mark, *et al.* (1996). A Review of Statewide Learner Competency and Assessment Systems. ERIC No. ED405482.

Lam, Tony, C. M. (1995). Fairness in Performance Assessment. *ERIC Digest.* On-line.

Mabe, P. A. III and West, S. G. (1982). Validity of Self-Evaluation Of Ability: A Review And Meta-analysis. *Journal of Applied Psychology*, 67: 280-296.

Montgomery, B. M. (1986). An Interactionist analysis of Small Groups Peer Assessment. *Small Group Behaviour*, 17: 19-37.

Morton, J. B. and Macbeth, W. A. A. G. (1977). Correlation Between Staff, Peer and Self-Assessments of Fourth-Year Students in Surgery. *Medical Education*, 11: 167-170.

Newkirk, T. (1984). Direction and Misdirection in Peer Response. *College Composition and Communication*, 35: 301-311.

Norton, Sue M. (1992). Peer Assessment of Performance and Ability: An Exploratory Meta-analysis of Statistical Artifacts and Contextual Moderators. *Journal of Business & Psychology,* 6(3): 387-399.

Nowack, K. M. (1992). Self-Assessment and Rater-Assessment as a Dimension of Management Development. *Human Resource Development Quarterly,* 3: 141-155.

Nuehring, E. M. and Pascone, A. B. (1986). Single-Subject Evaluation: A Tool for Quality Assurance. *Social Work*, 31: 359-365.

O'Grady, Martin J. (1991). Assessment of Prior Achievement /Assessment of Prior Learning: Issues of Assessment and Accreditation. *Vocational Aspect of Education*, 43: 183-189.

Oldfield, K., *et al.* (1995). Peer and Self-Assessment at the Tertiary Level - An Experiential Report. *Assessment and Evaluation in High Education*, 20: 125-132.

Ongley, Pat (1970). Objective Testing. *Education and Training,* 12: 470-471.

Orpen, C. (1982). Student Versus Lecturer Assessment of Learning: A Research Note. *Higher Education*, 11: 567-576.

Orsmond, Paul, *et al*. (1997). A Study in Self-Assessment: Tutor and Student Perceptions of Performance Criteria. *Assessment and Evaluation in Higher Education*, 22: 357-369.

Orsmond, Paul, *et al*. (1996). The Importance of Making Criteria in the Use of Peer Assessment. *Assessment and Evaluation in Higher Education*, 21: 239-250.

Page, G. C. and Fielding, D. W. (1980). Performance on PMP's and Performance in Practice: Are They Related? *Journal of Medical Education,* 55: 529-537.

Patterson, E. (1996). The Analysis of Application of Peer Assessment in Nurse Education, Like Beauty, Is In The Eye of the Beholder. *Nurse Education Today,* 16: 49-55.

Pond, K., *et al*. (1995). Peer Review: A Precursor to Peer Assessment. *Innovations in Education and Training International*, 32: 314:323.

Rada, R., *et al.* (1993). Collaborative Learning and the MUCH System. *Computer and Education*, 20: 225-233.

Rushton, C., *et al.* (1993). Peer Assessment in a Collaborative Hypermedia Environment: A Case Study. *Journal of Computer-Based Instruction*, 20: 75-80.

Schrader, M. A. and Westphal, R. (1976). *Reliability of Student Evaluations of Student Performance. Final Report*. Cleveland, WI: Lakeshore Technical Institute.

Shore, Lynn M., Tetrick, L. E. and Shore, Ted H. (1998). A Comparison of Self, Peer, and Assessor Evaluations of Managerial Potential. *Journal of Social Behavior And Personality*, 13(1): 85-101.

Sirotnik, K. A., *et al.* (1996). The Validity of Administrator Performance Assessment Systems: The ADI As A Case-in-Point. *Educational Administration Quarterly*, 32: 539-564.

Somervell, Hugh (1993). Issues in Assessment, Enterprise, and Higher Education: The Case for Self-, Peer and Collaborative Assessment. *Assessment and Evaluation in Higher Education*, 18: 221-233.

Swanson, D. (1990). Issues in Assessment of Practice skills in Medicine. *Professions Education Research Quarterly*, 12: 3-6.

Turner, C. and Record, A. L. (1981). Self-Concept Clarity: Evaluation of Self, Peer, and Professional Ratings. *Measurement and Evaluation in Guidance*, 14: 38-45.

Vogel, J., Conger, J. C. and Keane, S. P. (1985). Comparability of Peer-Assessment Measures: A Multitrait-Multimethod and Selection Analytic Approach. *Journal of Psychopathology and Behavioral Assessment*, 7(4): 385-396.

Williams, E. (1992). Student Attitudes Towards Approaches to Learning and Assessment. *Assessment and Evaluation in Higher Education*, 17: 45-58.

Wood, P. (1978). Student and Peer Ratings of College Teaching and Peer Ratings of Research and Service: Four Years of Departmental Evaluation. Presented at *Annual Meeting of the American Educational Research Association*.