



Archived at the Flinders Academic Commons:

<http://dspace.flinders.edu.au/dspace/>

This is the peer reviewed version of the following article:

McAllister, S.M., Lincoln, M., Ferguson, A. and McAllister, L. (2010). Issues in developing valid assessments of speech pathology students' performance in the workplace. *International Journal of Language and Communication Disorders*, 45(1) pp. 1-14.],

which has been published in final form at
DOI:10.3109/13682820902745461. This article may be
used for non-commercial purposes in accordance with
[Wiley Terms and Conditions for Self-Archiving](#)

The original publication is available by subscription at:
[http://onlinelibrary.wiley.com/journal/10.1111/
\(ISSN\)1460-6984](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1460-6984)

© 2010 Royal College of Speech & Language Therapists.
All rights reserved. **Please note** that any alterations made
during the publishing process may not appear in this version.

ISSUES IN DEVELOPING VALID ASSESSMENTS OF SPEECH PATHOLOGY STUDENTS' PERFORMANCE IN THE WORKPLACE

Valid assessment students' workplace performance

Keywords: Clinical Education, Competency Based Assessment, Workplace-based learning

Abstract

Background

Workplace based learning is a critical component of professional preparation in speech pathology. A validated assessment of this learning is seen to be 'the gold standard' but it is difficult to develop because of design and validation issues. These issues include the role and nature of judgement in assessment, challenges in measuring quality, and the relationship between assessment and learning. Valid assessment of workplace based performance needs to capture the development of competence over time and account for both occupation specific and generic competencies.

Aims

This paper reviews important conceptual issues in the design of valid and reliable workplace based assessments of competence including assessment content, process, impact on learning, measurement issues and validation strategies. The paper then goes on to share what has been learned about quality assessment and validation of a workplace based performance assessment using competency based ratings. The outcomes of a 4 year national development and validation of an assessment tool are described.

Methods

A literature review of issues in conceptualising, designing and validating workplace based assessments was conducted. Key factors to consider in the design of a new tool

were identified and built into the cycle of design, trialling and data analysis in the validation stages of the development process.

Main contribution

This paper provides an accessible overview of factors to consider in the design and validation of workplace based assessment tools. It presents strategies used in the development and national validation of a tool COMPASS™, used in every speech pathology program in Australia, New Zealand and Singapore. The paper also describes Rasch analysis, a model based statistical approach which is useful for establishing validity and reliability of assessment tools.

Conclusions

Through careful attention to conceptual and design issues in the development and trialling of workplace based assessments, it has been possible to develop the world's first valid and reliable national assessment tool for assessment of performance in speech pathology.

What this paper adds

What is already known on this subject: Valid and reliable assessment of workplace based learning is essential in professional entry programs but is often thwarted by difficulties in design and validation of workplace based assessment tools. These challenges have given rise to the belief that performance in the workplace cannot be validly quantified, particularly for high stakes examinations. As a result, most assessments of professional competence rely on de-contextualised assessment practices. However, it is believed that attention to conceptualisation and design issues and use of appropriate statistical analysis procedures of assessment data may overcome these obstacles.

What this study adds: This paper outlines conceptual issues that should be considered in assessment design such as measurement of performance, the nature of speech pathology competency and its developmental trajectory, assessment and its impact upon learning. Strategies to address these issues are illustrated through the description of the successful development and validation of COMPASS™, a competency based assessment of speech pathology student's performance in the workplace, now implemented in all speech pathology programs in Australian, New Zealand and Singapore

ISSUES IN DEVELOPING VALID ASSESSMENTS OF SPEECH PATHOLOGY STUDENTS' PERFORMANCE IN THE WORKPLACE

Introduction

All health professional prequalification programmes have at their heart the goal of ensuring that their participants, on graduation, will be fit for practice in their chosen discipline or speciality. Issues such as accountability to professions, clients and employers, describing and capturing the complexity of professional work, integrating notions of competency based/outcomes based education and a desire for assessment that identifies 'real' competence in the workplace are strong themes in the health education literature (Carraccio et al., 2002, Chapman, 1998, Cross et al., 2001, Higgs and Bithell, 2001, Norcini, 2005). There is considerable diversity in the way the notion of competence has been conceptualised and operationalised across different countries and different health professions, for example, see Wolf (1995) for a discussion of the integration of American competence-based approaches in the United Kingdom in response to government reforms. The outcomes of the research presented in this paper took place in the Australian context, in which a series of government initiatives placed the issues of conceptualising and operationalising competence at the forefront of both vocational education and migration policy (Hager & Gillis, 1995). Thus the speech pathology profession in Australia faced similar challenges and opportunities as experienced in the profession in the United Kingdom (RCSLT, 2007) and internationally in relation to how competence would fit within disciplinary education, qualifications, and ongoing professional development (Dawson, 1995, 2003; Ferguson, 2006).

The conceptualisation of notions of competence has tended to be implicit within different groups working within shared communities of practice, and so it is typical that when competence needs to be operationalised for assessment purposes that differences in approach become clear. These differences most notably are along the continuum from atomistic to holistic identification of competency, for example, the extent to which competence is considered to be behaviourally defined, observable, and able to be inferred. Ideally, this identification of competence would be carried out via a fair and defensible assessment process situated in the workplace with all its attendant complexities and compromises (Rethans et al., 2002). Speech pathology programs typically include a workplace assessment component based on clinical educators' ratings of student performance. As such these assessments are an important and integral part of speech pathology curricula. Only one example of psychometric validation of a speech pathology work place based assessment was found in the literature – the University of Western Ontario Clinical Grading System (Johnson and Shewan, 1988) which has some methodological limitations related to the issues discussed later in this paper. Development of validated workplace based assessments has been attempted by other professions, particularly medicine, but frequently discounted due to difficulties in ensuring that the assessment process is valid and reliable (Govaerts et al., 2006, Norcini, 2005). These difficulties have included concerns about identifying and managing sources of error such as rater effects, content specificity and the inherent variability of the workplace (Epstein and Hundert, 2002, McGaghie, 1993). These challenges have given rise to the belief that performance in the workplace cannot be validly quantified (Landon et al., 2003) and have resulted in continued reliance on de-contextualised assessment

practices such as Objective Structured Clinical Examinations (OSCE) with assumed, but unproven, links to real life professional performance (Spike et al., 2000). Nonetheless, validated assessments that measure students' ability to perform in the authentic workplace arena of professional practice continue to be seen as the 'gold standard' for assessment of whether a student is fit for practice (Wass et al., 2001). Thus calls continue for validated assessments to move from controlled educational environments to the unpredictably variable world of the workplace (Norcini, 2005).

This paper shares what has been learned about quality assessment and validation of a workplace based performance assessment using competency based ratings, that was developed during a 4 year research programme. This assessment format and process was validated via a national research process involving students from 7 Australian universities (full details can be found in McAllister, 2005). Rasch analysis, a model based statistical technique for validating assessments (Bond and Fox, 2007) was used in combination with classical test statistical approaches, and showed the assessment format to be reliable and valid (McAllister, 2005). The research led to the development of COMPASS™: Competency Based Assessment in Speech Pathology (McAllister et al., 2006), and is now integrated into all speech pathology curricula in Australia, New Zealand and Singapore (see Ferguson et al., 2008 and Lincoln et al., 2008). Programs are collaborating on using the interval level competency scores derived from COMPASS™ ratings to investigate the quality of curricula through internal and external benchmarking and research activities (Lincoln et al., 2008). Issues identified during the development of COMPASS™ included how to ensure that the assessment truly reflected the nature of speech pathology professional work, accurately measured the development of students' competency in

carrying out this work and supported students and clinical educators in their learning and teaching roles. The solutions arising from the COMPASS™ research that addressed these issues will be described.

Conceptual considerations in designing and validating performance assessment

Measurement processes: Ratings and the role of judgement in performance assessment

Assessment of competence through observation of real workplace performance results in a focus on the role of raters and ratees in this process and their respective impact on validity (Govaerts et al., 2006). The rater in performance assessment is characterised as a primary source of measurement error, which then limits the degree to which generalisation from assessment to performance can be assumed. In particular, concern is expressed over the role of judgement in determining whether students' performances are at a particular level and the subjective influences upon this judgement (Alexander, 1996). Reviews of the literature on rating performance suggest that there are multiple sources of error attributable to raters (Landy and Farr, 1980, Woehr and Huffcutt, 1994). Frequently rater training is cited as the solution to these issues.

However, not all research has found evidence to support this position. Generalisability studies have found that the rater or judge behaviour generally contributed less to error variance than other factors such as ratee knowledge, tasks sampled (Govaerts et al., 2002, Shavelson et al., 1993) and case specificity. Thus other aspects of assessment content, design, or process may have a greater impact on assessment validity than rater behaviour.

To a certain degree the point is moot, in that the aim of achieving totally scientific and

objective assessment of human behaviour is an unachievable ideal (Bitzer, 1999), and it can be argued that all assessment of learning and performance is inherently subjective (Leach et al., 2001). In fact, judgement is both inevitable and integral to the design of any valid assessment tool (American Educational Research Association, 1999, Schuwirth and van der Vleuten, 2006). Thus the primary issue becomes: how to support quality judgement in assessment?

First, assessment tools need to be efficient for use with the large majority of students and highly effective for the minority about whom concerns are held, which results in calls for both brevity and detail (Hunt, 1992, McAllister, 2005). Assessment formats need to include features to ensure that judgement is supported through provision of a rich source of information and context to guide the assessor's judgement (Jones, 2000). In addition, development of scoring rubrics that focus on observed qualities of performance can improve the psychometric qualities of an assessment (Wolfe and Gitomer, 2001). At the same time it is important to allow for the exercise of professional judgement and not over specify the competencies or outcomes of interest by resorting to checklists which trivialise and neglect the holistic nature of professional competency (Jones, 2000, Norman et al., 1991). Using global judgements that focus on the holistic nature of competency have been found to improve reliability and validity in ratings based assessments of performance in health education and to identify important aspects of performance such as transfer of competence (Cox, 2000, Epstein and Hundert, 2002, Keen et al., 2003).

Second, quality judgement can not be exercised in the absence of clearly defined competencies and without clear definitions of threshold standards for judging competence

thus the content of the assessment format is critical (Ilott and Murphy, 1997). This also includes ensuring that competencies that clinical educators have concerns over rating, due to their perceived subjectivity such as attitudes and values (Duke, 1996), are carefully described in assessment documentation.

Finally, it is important to ensure that the assessment is based on evidence of sufficient quantity and quality by ensuring thorough formative assessment occurs with ongoing feedback and consideration of performance over the workplace placement (Peters et al., 2001). This should include observing students working with as wide a variety of cases as possible to avoid the pitfall of case specificity that has been consistently identified as having a significant negative impact upon the validity of OSCE assessments (Newble et al., 2000).

Quality measurement

Assessment is essentially an endeavour to measure students' behaviour as a basis for drawing conclusions, in this case, as to whether a student is sufficiently competent to enter the profession. Educational researchers identified concerns regarding the epistemological principles used to validate assessment tools in the early 1990s (Friedman and Mennin, 1991, Moss, 1992) and these concerns persist to this day (Schuwirth and van der Vleuten, 2006). In particular, there are concerns that the reductionist outcomes of only recognising learning as learning which can be measured by assessments designed using traditional or classical test theory psychometrics, neglect other qualitative or holistic aspects of learning. For example, focussing on traditional notions of reliability results in assessment designs involving highly controlled testing environments does not mirror work environments of graduate health professionals. Therefore choice of

measurement model used to validate an assessment will reflect the developer's understanding of the nature of valid measurement of learning. The measurement model will in turn affect the type and quality of information available to guide decision making regarding assessment content and process. Therefore any discussion of issues in developing valid assessments must include consideration of the statistical strategies used to guide design and validate the tool to ensure that it measures competency accurately.

There are two major approaches used to validate assessment tools, Classical Test Theory (CTT) and Modern Test Theory. Modern Test Theory (also known as Item Response Theory) includes a number of statistical approaches, one of which is Rasch analysis (Embretson & Reise, 2000). Modern Test Theory approaches involve comparing data generated by an assessment tool to a model of how data generated by a high quality measurement (assessment) tool should behave. Rasch differs from the majority of Item Response Theory approaches in that it tolerates some degree of chance variation in performances on an assessment tool and considers variation in human performance to be normal and expected (Bond and Fox, 2007). This is also in contrast to CTT approaches that assume errors are normally and uniformly distributed in persons, have an expected value of zero, and are uncorrelated with other variables (Embretson, 1999). Both Rasch and CTT aim to assess a latent variable (i.e. inferred characteristic of interest such as 'competence') and relate it to performance on the test items (Embretson, 1999). However there are fundamental differences in their approach and utility for evaluating the validity of a performance assessment tool.

Using Rasch analysis allows for a closer adherence to fundamental principles of measurement than CTT while attempting to quantify latent traits of human beings.

Measurement involves first proving that the attribute being measured is quantitative and second, constructing procedures for numerically estimating the magnitudes of the attribute being measured (Michell, 1997). Michell suggests that current practice in the measurement of non physical phenomena such as human behaviour has ignored these basic requirements by simply assigning numbers to observed phenomena according to a rule (Michell, 1997). This misconception regarding measurement of human behaviour has resulted in test developers treating the numbers assigned to behaviours via an assessment tool e.g. derived from rating scales, as if they were interval or ratio data. For example, raters are provided with a scale with a predetermined number of categories to use to represent a student's performance. It is assumed that all these categories of discrimination are in fact possible, are of equal size and are interval data possessing additive properties that allow them to be validly summed (Zhu, 1996). All of these assumptions warrant examination before proceeding to conduct mathematical processes such as parametric statistical analysis on any kind of rating scale data (Bond and Fox, 2007, Linacre, 2002). For example, data from rating scales are ordinal in nature, representing 'more' of a quality but exactly how much 'more' of a quality the observed change in behaviour represents is not quantified. Therefore such data should not be subjected to parametric statistical processes, such as those used by CTT, that assume the data is interval (Zhu, 1996).

In contrast, Rasch modelling aims to develop measures that are quantitative in nature by identifying how many rating categories actually function in the data generated by the assessment, quantifying the size of the steps between rating categories and identifying equal increases in magnitude in the attribute being measured. Once this process is

completed, the rating scale data can then be converted into meaningful, quantifiable rating categories and the item and person measures examined. This occurs through summing the ratings for each item, as derived from the validated rating scale categories, to yield a raw score. A probabilistic algorithm that is applied to these scores to convert them into an interval measure (logit) that quantifies the degree to which the latent variable possessed by each person as measured by the assessment. Statistics are generated that identify how much error is present in the data and describe how much confidence can be held regarding the assessment tool as a device to measure the amount of the latent variable each person possesses (Bond and Fox, 2007; McAllister, 2008). These logit measures can be validly examined via parametric statistics to identify whether any probable relationships exist between the quantified amount of competency and other variables of relevance.

The process of analysis is guided by comparing the data generated to the Rasch model. This model assumes that, if the assessment items are effectively sampling a unidimensional trait that is quantifiable, the data will satisfactorily meet the requirements of the model. The requirements are logical and simple; if an ordered continuum in item difficulty and assessee performance exists and the assessment measures it effectively, the following will apply: some items will be more difficult to rate highly on than others; more competent students will tend to rate higher than less competent students (albeit with a reasonable amount of variation in performance); and less competent students will be more likely to have lower ratings on more difficult items. If the data from the performance ratings does not conform to this model, statistical information is provided to guide the examination of variations from this pattern and revision of the measurement

instrument (assessment) so that it functions more effectively as a measurement tool i.e. generates data that behaves in the way predicted by the model.

On the other hand, the statistics used in CTT are descriptions of the raw data and not measures (Bond and Fox, 2007) and some are not entirely able to meet the requirements of the task for which they are used (Clark and Watson, 1995). In CTT measures are confounded with the sample of respondents as the difficulty of an item is defined as the proportion of respondents passing the item, such that an item may appear easier if the sample happened to include more people who could perform highly on the item and vice versa. Therefore item difficulty depends on the ability of the sample on which it is being used, which hopefully is representative of the group for which the test is intended (Barnard, 1999, Bond and Fox, 2007, Embretson, 1999). Rasch modelling is able to estimate item difficulties independently of the distribution of abilities of the particular group the assessment has been performed on, often termed 'person free' estimates (Bond and Fox, 2007). This means that estimation of item difficulties is not affected by the range of abilities of the particular sample used, assuming the items in the assessment are appropriate for that sample (Bond and Fox, 2007). This has the added advantage of Rasch analysis resulting in reliable estimates of item difficulty and student performance based on smaller samples than CTT approaches (Linacre, 1994).

In summary, Rasch analysis allows assessment developers to move from raw observations with unclear measurement properties to well-defined, abstract linear measures with realistic estimates of precision and explicit quality control (Wright, 1999).

Assessment and learning

Assessment not only serves a regulatory or gate keeping purpose to ensure that graduates can perform ‘on the job’ but it also fundamentally drives learning by communicating to the assessees what is important and valued, or ‘what counts’ (Wass et al., 2001). Thus the content and processes of the assessment must ensure that students’ attention is drawn to competencies that are congruent with what the profession considers important, and have positive impact on the students’ learning. Boud (2000) cautions that the focus on summative assessment for the purposes of certification, grading and public measures of performance and accountability relegates the focus on learning to the background and so negatively affects learning while at the same time attempting to measure it.

Formative assessment, through provision of feedback and opportunities to remediate or further develop performance, has an important role in providing students an opportunity to develop and practice lifelong learning skills such as self evaluation, problem solving, and strategies for the acquisition of new skills and knowledge (Boud, 2000). As formative assessment is student focussed, it is an opportunity to facilitate the student’s active involvement in and responsibility for their own learning and may enable them to engage as adult learners in their own learning, supporting the development of the critical professional competency of lifelong learning (Boud and Falchikov, 2006, Leach et al., 2001). In addition, it provides students with the opportunity to identify whether their perception of their performance is similar to that of others, thus developing their skills and confidence in their ability to self-monitor (Robertson et al., 1997). Finally, the student is more likely to internalise the learning targets, set his/her own goals in relation to these and self monitor their progress towards them (Boud, 2000, Brookhart, 2001).

These are critical outcomes if we wish to ensure that speech pathologists are competent to practice at graduation and on into the future.

The final aspect of assessment to consider is that of the authenticity or validity of the assessment tasks (Hays et al., 2002, Linn et al., 1991), which is closely linked to the previous discussion regarding the nature of the rating task. There is general recognition that the most authentic assessment will be one that assesses the real task of functioning effectively as a speech pathologist in the workplace. However, attaining this ideal has been frequently frustrated by difficulties in applying the predominant CTT measurement and assessment model to performance based workplace assessment. This includes issues previously mentioned such as controlling testing environments to safeguard reliability resulting, and assumptions that reducing professional practice in smaller, simpler and assessable units of behaviour (e.g. performing an oral examination) will automatically generalise to the more complex and integrated performances required in the workplace (Friedman and Mennin, 1991).

Assessment content

Quality observation and assessment relies on clear identification of what constitutes professional competence and what behaviours will indicate a competent professional performance is evident. This includes both the types of competencies and description of a developmental continuum.

Occupational and Generic Professional Competencies

Detailed and narrow descriptions of professional competency trivialise the nature of competent professional performance through reducing it to minutely detailed descriptions of observable behaviours related to specific occupational tasks (Hager et al., 1994). This

reductionist approach to describing professional work neglects those aspects of our work that are the result of holistic integration and coordination of these specific occupational tasks (Hager et al., 1994). For example, in the recently developed framework for competencies of newly qualified practitioners, the Royal College of Speech and Language Therapists, includes dimensions of competence that are generic in nature (Communication; Personal and People Development), fundamental to disciplinary practice (Assessment and Care Planning to meet Health & Well-being Needs; Health & Well-being Intervention) and specific to workplace demands (Health, Safety & Security; Service Improvement; Locally-driven Competencies) – RCSLT, 2007. Speech pathology in Australia has had a statement of occupational competencies, the Competency Based Occupational Standards (CBOS) document since 1993, which is subject to an ongoing process of revision and updating (SPAA, 2001). This document outlines 7 occupational competencies that describe in global terms what speech pathologists do: Assessment; Analysis and Interpretation; Planning of Speech Pathology Intervention; Speech Pathology Intervention; Planning, Maintaining, and Delivering Speech Pathology Services; Professional, Group, and Community Education; and Professional Development. While the CBOS descriptions of competency have avoided narrow, detailed descriptions of speech pathology work, the holistic and integrated nature of competent professional practice is embedded in the detail of the document and therefore not explicit.

The need to make this holistic and integrated aspect of competence explicit has been identified by students, field educators, and academics (Epstein and Hundert, 2002, McAllister, 2005), This aspect appears to represent the ‘something else’ of competent

professional performance, or those aspects of professional practice that are more than the demonstration of particular occupational competencies, where the 'whole' is greater than the sum of the parts. This conceptualisation of holistic workplace competency can be understood as arising from a combination of occupational competencies and generic (or key) competencies. The generic competencies support the development and holistic integration of occupational competencies across work tasks both in the present and on into the future (Harris et al., 1995, McAllister, 2005). The nature of these generic competencies, that enable the competent performance of the occupational competencies, has been frequently raised in discussions regarding competent professional practice and has included aspects such as reasoning, lifelong learning, and professionalism, to name a few (e.g. Rose & Best, 2005).

Competent performance must be characterised by behaviours that indicate that a speech pathology student is able to engage in appropriate professional action, not just at the point of final assessment prior to entering the profession, but continuing on into his/her professional life. Ongoing development and maintenance of competence relies on competent exercise of complex professional judgement and action (Hager, 2000). This exercise of professional judgement and action is the result of integrated combinations of knowledge, skills, and personal qualities (Carter, 1985, McAllister, 2005). Professional judgement must be exercised across all tasks and contexts of the speech pathology profession. It includes performing occupational competencies as well as engaging in behaviours indicative of generic professional competencies that result in a holistic action based on integrated performance of occupational competencies. These generic competencies are considered to be instrumental in the development, maintenance and

transferability of occupational competencies over a speech pathologist's working life (McAllister, 2005).

Development of Competence

Medical and other health professional literature clearly identify that competence is developmental (Benner et al., 1996, Dreyfus and Dreyfus, 1996). However, the nature and dimensions of this development are not well described, with workplace assessments of competency generally requiring a clinical educator to rate a student against a poorly described range of performance levels or parameters. For example, Wilkinson and Frampton (2003) reported on use of a performance assessment where trainee interns were rated on a 7 point scale ranging from a rating of 1 indicating that the ratee is “clearly incompetent on this item” through to 7 “Extremely competent”. However, what constitutes clear incompetence or extreme competence, or any of the 5 other degrees between these points was not identified. This does not meet the requirement for scoring rubrics to clearly identify behaviourally observable changes in quality of performance (Wolfe and Gitomer, 2001).

While behaviourally anchored rating scales are not necessarily superior to other types of rating scales (Fay and Latham, 1982, Gomez-Mejia, 1988, Kingstrom and Bass, 1981), Landy and Farr (1980) found in their extensive literature review that scale anchors are important. Their review also indicated that behavioural anchors may be better than numerical or adjectival ones. This advantage is suggested to be due to behavioural descriptors ensuring that a rater has a clear understanding of the rating task and rigorously developed anchors that are more than simple descriptive labels such as poor/average/excellent, support their judgement more effectively.

It seems likely that if the progressive nature of the development of competency is not identified, raters and ratees will need to rely on their own perceptions or concepts of what constitutes progressive performance, as Govaerts et al (2006) identifies – raters are not value free measurement instruments and ratees can also influence the process. This may introduce uncertainty for all parties as to what exactly is being rated and what would represent improvement in performance. Not only does this introduce the possibility of idiosyncratic variance between raters, but also neglects the formative aspects of assessment as criteria for improved performance are not explicit.

In summary, four major conceptual issues in designing assessments of performance in the workplace have been described, and guided subsequent development of COMPASS™. These include

1. Measurement processes – workplace assessments where the student's performance is observed, will rely on ratings by observers. Quality professional judgement can be supported by providing a rich source of information and context to inform this judgement, in combination with clearly defined competencies and threshold standards. Furthermore, these judgements should be exercised in the light of sufficient quantity and quality of observational evidence.
2. Quality measurement – the measurement model underpinning the statistical validation of the assessment tool should be carefully scrutinised to ensure that it provides information that will effectively guide the design of the assessment tool. Rasch measurement, rather than the more commonly used Classical Test Theory approaches, was identified as having the greatest utility for validating

a performance assessment.

3. Assessment and learning – given that assessment drives learning, it is critical to ensure that assessment content is authentic and relevant, and formative assessment included to ensure a focus on learning is maintained and to facilitate the development of the professional skill of lifelong learning.
4. Assessment content – both the performance of occupational tasks (occupational competencies) and the ability to develop and holistically integrate the ability to perform these tasks across the context of speech pathology professional practice should be included in an assessment of workplace competency. The developmental trajectory from starting as a novice in a first workplace placement through to ready to enter the profession should be identified so that information about the nature of the learning process guides both the student and clinical educator.

COMPASS™: Competency assessment in speech pathology

The initial process and development of COMPASS involved two phases undertaken over a three-year period: Development of the tool content and design, and a field trial of the prototype assessment. The development phase involved a reiterative process incorporating multiple sources of evidence (including previous research, consultation, qualitative focus group research). The field trial phase was conducted from February to November 2003, involving seven Australian universities, 219 different students and 88 different clinical educators. The following description has been organised so

that information drawn from across these phases is grouped in relation to each component of the new assessment tool.

Overview of Design

The design of COMPASS™ was informed by the conceptual considerations described above, and subsequently validated through an Australia wide field trial using a combination of Rasch analysis, classical statistics and qualitative data. COMPASS™ is comprised of four elements: Assessment Booklet, Assessment Resource Manual, Technical Manual, and Training Modules.

Assessment Booklet

The assessment booklet comprises a set of ratings to be made by the clinical educator for each competency based on his/her observations of the student over the placement. A detailed set of formative only ratings is carried out half way through the placement (regardless of length), with a brief set of summative ratings made at the end of the placement. These ratings are represented on a Visual Analogue Scale (VAS) (see figure 1) and are guided by behavioural descriptors that are provided on a sheet that opens to the side of the booklet and is always visible while rating. The same set of behavioural descriptors is applied to each competency, and the competencies are written so that they can be applied to any client group or service delivery model. Clinical educators are advised to ask students to self rate and to collaboratively determine the final ratings for both the formative and summative assessments.

Insert figure 1 here

Assessment Resource Manual

This booklet supplies supporting layers of detail that the clinical educator and student can access when making more specific and complex decisions about the student's competency, or to plan for learning across the placement. It includes more detail on how to conduct ratings, information on terminology, a detailed analysis of Competencies and the Behavioural Descriptors. In addition, there is an extensive section where examples of behaviours that would be representative of different levels of performance are described for each of the competencies.

Technical Manual

This manual is used by university programs to score the COMPASS™ ratings, and is not made available to students or clinical educators to ensure that the validity of the tool is not degraded.

Training Modules

A set of 3 training modules are provided that can be conducted with a group or as a self study package. These modules use a frame of reference approach (Woehr and Huffcutt, 1994) to promote clinical educators' confidence in using the tool for both assessment and teaching.

Specific design elements

Competencies

The assessment includes 11 competencies comprising the 7 occupational competencies already identified by the Australian speech pathology profession in CBOS. Four other generic competencies (as discussed above) were identified during the research based on student and clinical educator opinion, in combination with the literature (McAllister,

2005; McAllister et al., 2008). These were classified into four major themes or units of competency: reasoning, lifelong learning, communication, and professionalism (McAllister, 2005). A detailed description of each of these competencies was developed to parallel the CBOS format for consistency, with each unit of competency being subdivided into related elements and elaborated with performance indicators and cues (see table 1 for definitions). Table 2 outlines the unit and element levels of these competencies. The validation process confirmed that all 11 competencies (4 Generic Professional and the 7 Occupational Competencies described in the CBOS) acted in concert to holistically capture the nature of speech pathology practice, confirming that an approach that captures the interacting generic and occupational aspects of competency is appropriate for assessment of speech pathology practice.

Insert Tables 1 and 2 here

Visual analogue scale

A VAS was chosen for two reasons. Firstly, rather than assuming clinical educators could make a certain number of performance discriminations, the VAS was analysed using Rasch Rating Scale Analysis Model (Bond and Fox, 2007). This identified that clinical educators could reliably discriminate 7 different and equal levels of student performance. The ratings are converted into one of 7 categories or levels of performance based on this analysis for each of the 11 competencies rated for the summative assessment. As these categories represent equal sized increments in performance, they can be summed and converted into a competency score by the university program with the assistance of the Technical Manual (McAllister, et al., 2006). The VAS format was retained from the format trialled during the research, rather than converted into a 7 category scale as

feedback indicated that it was generally well received by both students and clinical educators (McAllister, 2005; McAllister et al., 2004). This format also enabled clinical educators to make judgements and represent progress without needing to consider whether the rating mark represents a change in category of performance.

Behavioural Descriptors

Ratings on the VAS are carried out on the basis of the clinical educator's professional judgement of the student's performance in relation to each of the competencies with reference to the 3 Behavioural Descriptors. These behavioural descriptors were developed after consideration of the literature and current practice in practicum assessment, and map progressive development in the student's ability to

- Manage complexity as described by the Structure of Observed Learning Outcomes taxonomy (Biggs and Collis, 1982) that describes increasing abilities to identify and relate relevant sources of evidence.
- Integrate knowledge and develop expertise through experience and transform how this knowledge is used in clinical situations to inform judgement, as described by Benner (1984, 1996) and Dreyfus and Dreyfus (1996). This incorporates notions of 'bottom up' and 'top down' reasoning described by Govaerts, van der Vleuten et al (2006).
- Competently perform with decreasing degrees of support/guidance, related to theories espoused by Bruner (1983), Anderson (1988) and Brasseur (1989).

These concepts were used to develop a description of 3 levels of performance (novice, intermediate, entry-level) which clinical educators were asked to use to guide their rating decisions. See table 3 for the descriptions of these levels. These descriptors are displayed

on a flap that is folded out to the side of the booklet so they can be referred to at all times while rating. They are also described in further detail in the Assessment Resource Manual along with detailed example of what kinds of behaviours would be indicative of each of the three levels of performance for each competency (see table 4). The VAS is labelled with each of the 3 levels (see figure 1).

Entry-level was chosen as the behavioural anchor for the end of the scale (with the option to tick ‘above entry level’), because a competency based assessment is focussed on ensuring that students are sufficiently competent to enter the workforce. This level was matched with expectations in the CBOS (SPAA, 2001), which lined up well with the description of ‘competent’ (Benner, 1984, Dreyfus and Dreyfus, 1996). The validation procedure identified that, in actuality, speech pathology educators were able to identify 4 other distinct degrees of performance on this continuum (7 in total), with further research required to develop descriptions of these categories of performance. Feedback from students and clinical educators indicated that this description of the development of competency matched their perception of the developmental task at hand (McAllister et al., 2004).

Insert tables 3 & 4 here

Relationship of design to conceptual considerations for design of

COMPASS™

The design of COMPASS™ accepts that judgement plays a role in assessment and that these judgements can be effectively represented by ratings. COMPASS™ supports quality judgment processes in a number of ways. A rich source of information and

context to guide judgement is provided in the Assessment Resource Manual. Judgement is supported through developing a balance between providing clearly defined and relevant competencies but not being over prescriptive. The taxonomy adopted by CBOS and applied to the Generic Professional Competencies supports this process by making performance based statements about the area of competence (Unit Level) and the major areas of performance included under each Unit (Elements), and a number of clarifying examples (Performance Indicators and Cues) (see table 1). The Assessment Resource Manual further supports this in an assessment context by providing specific examples of behaviours that may be observed that relate to each competency, with examples given for each level of performance (Novice, Intermediate, and Entry-Level) (see table 4). The nature of required performances could be further tailored by the clinical educator for their workplace so examples refer to the client group or services provided in which the student will be expected to demonstrated developing competence. These strategies provide students and clinical educators with 'layers' of detail that can be accessed as required e.g. to set specific learning goals after the mid placement formative assessment process or to provide explicit guidance to a student whose performance is of concern.

The COMPASS™ assessment process requires the student to be rated by the clinical educator who is working directly with the student over the duration of the placement, as opposed to an external assessor. Therefore the student is assessed by the clinical educator who has had multiple opportunities to observe the student's performance, across multiple clients and contexts, as well as having engaged in both the detailed formative assessment and the briefer summative assessment. The assessment design requires, and is validated on the basis of, a formative assessment being carried out half way through the student's

placement, regardless of length. This ensures that the student's learning is promoted through having the opportunity of engaging in a low risk, non summative assessment process that is focussed on their learning. This also ensures the judgement of the student's competence is based on sufficient quality and quantity of evidence, that has been reviewed on more than one occasion (i.e. mid placement) and all of which has been gathered in the authentic environment of the workplace. This workplace focus promotes a highly relevant assessment experience for the student and, in combination with the formative assessment process, aims to retain a strong focus on the learning occurring throughout the placement and as a result of the assessment. It is also recommended in the COMPASS™ instructions that all assessment is carried out as a negotiated joint process with the student, thus providing the student with guided opportunities to develop and practice lifelong learning skills. This process also encourages the clinical educator to make assessment decisions explicit and evidence based.

This relevance is further supported by the assessment content that, as mentioned previously, was found to represent a unidimensional construct of competency where the CBOS (Occupational Competencies) and Generic Professional Competencies act in concert to generate competent performance in the workplace. Including competencies that are well accepted by both students and clinical educators (McAllister et al, 2004) and reflect their understanding of speech pathology competencies ensures that meaningful learning is supported. While the implications of this form of assessment for speech pathology academic curricula is beyond the scope of the present paper, it can be noted that validation of competency assessment provides a strong foundation for curriculum development, through the clear specification and communication of graduate attributes

and outcomes (Barrie, 2006). This form of assessment also has the potential to establish an explicit nexus between classroom and workplace learning that fosters the development of clinically relevant teaching practices in the classroom.

As discussed previously, the Rasch model was chosen to validate COMPASS™ and provide the basis for the scoring process provided by the Technical Manual. This validation (as described earlier) ensured that a rating scale was constructed that enabled developmental progression to be effectively represented. The validation process provided evidence that the behavioural descriptors were effective in guiding the rating process. These ratings could then be converted into interval level data, validly summed and a total competency score derived that can then be appropriately subjected to parametric statistics. This feature of the data generated by COMPASS™ is now being explored by Australian and New Zealand Speech Pathology programs to identify its utility for tracking and evaluating progress of both individuals and cohorts of students, evaluate curriculum and conduct research.

Summary

This paper outlines the issues in developing valid competency based assessments of students' performance in the workplace and describes COMPASS™, a solution developed and validated by the Australian speech pathology profession. COMPASS™ is an assessment resource that values the expertise of clinical educators in judging the competence of learners, is based on sound measurement principles, and ensures that the learning process is supported and not subverted by the validation or assessment process required to gather of evidence of achievement (Schuwirth and van der Vleuten, 2006). The speech pathology profession in Australia is committed to the ongoing development

of this tool, and using the competency measures generated to develop evidence on which to base decisions regarding pedagogical practice.

Tables

Table 1 Definitions of competency taxonomy*

Taxonomy Level	Definition
Units	Heading and descriptive paragraph outlining a broad area of professional activity.
Elements	Specific activities carried out within the specified unit.
Performance criteria	Example statements of behaviours that enable you to infer that the student is carrying out the elements of competency to an acceptable standard.
Cues	Examples of evidence of observable behaviours that assure you the student has met the performance criteria. Can include demonstrated knowledge(s), skills, personal qualities, practical considerations and other relevant contextual information.

*adapted from CBOS, and Hager et al (1994)

Table 2 Generic Competencies: Units and Elements only*

Unit of Generic Competency
1. Clinical Reasoning
1.1 Uses effective thinking skills to ensure quality professional practice
1.2 Integrates collaborative and holistic viewpoints into professional reasoning
1.3 Uses sound professional reasoning strategies to assist planning for all aspects of service management
2. Professional Communication
2.1 Uses interpersonal communication skills to facilitate effective professional practice
2.2 Uses oral and written reporting and presentation skills to successfully meet speech pathology objectives
2.3 Communicates effectively with work teams
3. Lifelong Learning
3.1 Reflects on performance
3.2 Structures own learning/professional development
3.3 Demonstrates an appropriate attitude to learning
3.4 Able to change performance
4. Professional behaviour
4.1 Displays effective organisational skills
4.2 Conducts self in a professional manner
4.3 Discharges administrative responsibilities effectively
4.4 Possesses a professional attitude/orientation
4.5 Demonstrates ethical behaviour

* see McAllister (2005) and COMPASS™ (McAllister et al, 2006)

Table 3 Behaviours describing developmental continuum of competency*

Developmental Level	Description
---------------------	-------------

Novice Student	<p>High degree of supervisory support Can recall some aspects of relevant theory Needs support to:</p> <ul style="list-style-type: none"> • Draw conclusions about a client • Develop a plan for action • Understand the total clinical situation • Apply problem solving strategies, principles and theory. <p>Spends a high degree of time and effort in meeting clinical responsibilities Highly focussed on own performance rather than the client.</p>
Intermediate	<p>The complexity of the client, the workplace environment and the student's previous experience determines:</p> <ul style="list-style-type: none"> • Degree of supervision (moderate to low) • Ability to recognise the meaningful aspects of a situation. <p>Recognises several aspects of a problem but not all, and relates these to the client's needs and is able to:</p> <ul style="list-style-type: none"> • Draw some accurate conclusions about a client • Develop some plans for action • Recognise some important aspects of the total clinical situation. <p>Requires support to:</p> <ul style="list-style-type: none"> • Recognise and prioritise all aspects of a situation • Flexibly apply problem solving strategies, principles and theory. <p>Developing automaticity resulting in:</p> <ul style="list-style-type: none"> • A moderate expenditure of time and effort • Greater ability to focus on the situation than on own performance • A developing ability to use observation to assist clinical reasoning.
Entry-Level (Competent)	<p>Performs the majority of his/her work independently and competently Seeks support if the situation is new or a number of features about the client or workplace setting combine to create complexity Identifies meaningful aspects of problems and integrate these to generate a number of logically possible conclusions. Conclusions/actions will be modified with new information Prioritises appropriately</p>

Sufficiently automatic and maintains a focus on the client or situation
 Carry out his/her work in an efficient and timely manner.

*extracted from COMPASS™ (McAllister et al., 2006) and reproduced with permission from Speech Pathology Australia

Table 4 Examples of observable behaviours indicating performance levels on COMPASS™ Occupational Competency of ‘Assessment’*

Level of Performance	Examples of observable behaviours
Novice	<p>Identifies the presenting condition, related issues and the significant people in the client’s life with a high level of input from the clinical educator.</p> <p>Develops an adequate assessment plan with a high level of support from the clinical educator including discussion, reference to theory and direct feedback from the clinical educator.</p> <p>Inconsistent in his/her ability to correctly perform formal assessments, may need to perform these collaboratively.</p> <p>Is strongly focussed on the process of administering the assessment and may not recognise qualitative information that can be gained through observation of the client’s response to the assessment situation.</p>
Intermediate	<p>Partially establishes the nature of the presenting communication and/or swallowing condition and issues. Identifies most of the significant other people in the client’s life. Will need support from the clinical educator to ensure a comprehensive picture is developed, and direction to ensure that all relevant information is accessed and collated.</p> <p>Needs time to develop an assessment plan, with opportunity for discussion, reference to theory and reflection on feedback from the clinical educator.</p> <p>Consistently performs formal assessments but is still focussed on the process and neglects qualitative information that can be gained</p>
Entry Level	<p>Independently establishes and documents the presenting condition and issues, identifies the significant people in the client’s life and collates information on the client for familiar conditions.</p> <p>Seeks and requires support through consultation, collaboration or supervision for situations they have not previously experienced, or where a number of features of the client and/or</p>

context combine to create complexity.

Develops a comprehensive and appropriate assessment plan, and seeks and requires input to finalise the plan for more complex clients and/or situations.

Identifies formal assessment booklet that are appropriate for use and conducts formal assessments in a way that ensures the validity and reliability of the test results.

Develops and carry out informal assessments

*extracted from COMPASS™ (McAllister et al., 2006) and reproduced with permission from Speech Pathology Australia

Figures

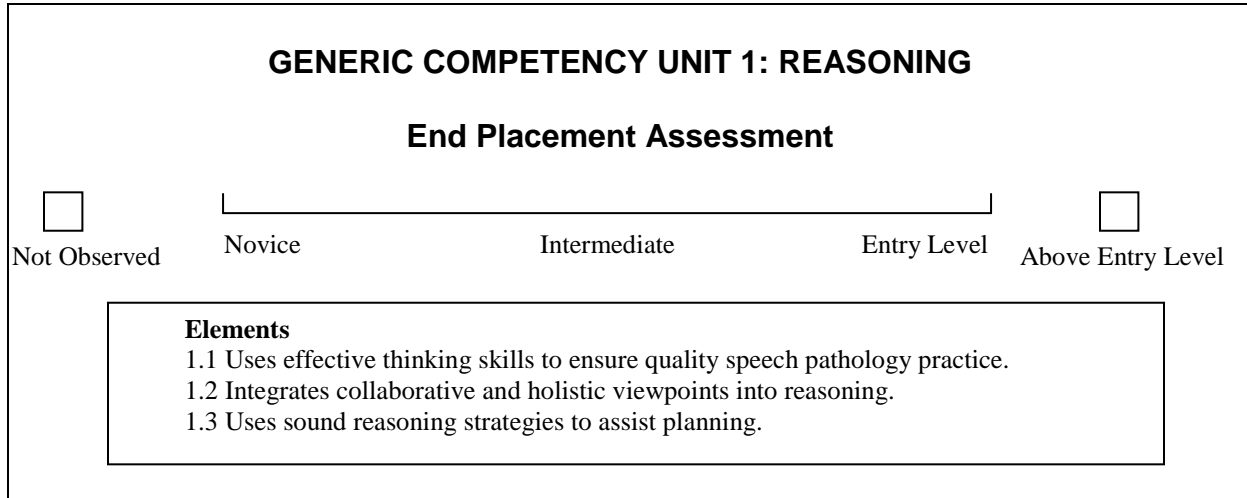


Figure 1. Example of COMPASS™ Rating Scale*

*extracted from COMPASS™ (McAllister et al., 2006) and reproduced with permission from Speech Pathology Australia

References

- AERA (1999) Validity. *Standards for Educational and Psychological Testing*. Washington, American Educational Research Association.
- ALEXANDER, H. (1996) Physiotherapy student clinical education: The influence of subjective judgements on observational assessment. *Assessment & Evaluation in Higher Education*, 21, 357-366.
- ANDERSON, J. (1988) *The Supervisory Process in Speech-Language Pathology and Audiology*, Boston, College Hill.
- BARNARD, J. J. (1999) Item analysis in test construction. IN MASTERS, G. N. & KEEVES, J. P. (Eds.) *Advances in Measurement in Educational Research and Assessment*. Oxford, Elsevier Science Pty Ltd.
- BARRIE, S.C. (2006) Understanding what we mean by the generic attributes of graduates. *Higher Education*, 51, 215-41.
- BENNER, P. (1984) *From Novice to Expert: Excellence and power in clinical nursing practice*, New Jersey, Prentice-Hall.
- BENNER, P. A., TANNER, C. A. & CHESLA, C. A. (1996) *Expertise in Nursing Practice: Caring, Clinical Judgment, and Ethics.*, New York, Springer Publishing Company.
- BIGGS, J. B. & COLLIS, K. F. (1982) *Evaluating the Quality of Learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*, New York, Academic Press.
- BITZER, E. (1999) Assessing Learning in the MPhil (Higher Education Studies): Paving new ways with cornerstones. *Higher Education Research and Development Society of Australasia Annual International Conference*. Melbourne.
- BOND, T. G. & FOX, C. M. (2007) *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Mahwah, N.J., Lawrence Erlbaum Associates, Inc.
- BOUD, D. (2000) Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22, 151-167.
- BOUD, D. & FALCHIKOV, N. (2006) Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education*, 31, 399-413.
- BRASSEUR, J. (1989) The supervisory process: A continuum perspective. *Language, Speech and Hearing Services in Schools*, 20, 274-295.
- BROOKHART, S. M. (2001) Successful Students' Formative and Summative Uses of Assessment Information. *Assessment in Education*, 8, 153 - 168.
- BRUNER, J. (1983) *Child's Talk: Learning to use language*, New York, W. W. Norton & Company.
- CARRACCIO, C., WOLFSTHAL, S. D., ENGLANDER, R., FERENTZ, K. & MARTIN, C. (2002) Shifting paradigms: From Flexner to competencies. *Academic Medicine*, 77, 361-367.
- CARTER, R. (1985) A taxonomy of objectives for professional education. *Studies in Higher Education* 10(2): 135-149.
- CHAPMAN, J. (1998) Agonising about assessment. IN FISH, D. & COLES, C. (Eds.) *Developing Professional Judgement in Health Care: Learning through the critical appreciation of practice*. Oxford, Butterworth-Heinemann.

- CLARK, L. A. & WATSON, D. (1995) Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.
- COX, K. (2000) Examining and recording clinical performance: A critique and some recommendations. *Education for Health*, 13, 45-52.
- CROSS, V., HICKS, C. & BARWELL, F. (2001) Exploring the gap between evidence and judgement: Using video vignettes for practice-based assessment of physiotherapy undergraduates. *Assessment & Evaluation in Higher Education*, 26, 189-212.
- DAWSON, V. (1995). C-BOS for students: Developing professional competencies in speech pathology students (part 1). *Australian Communication Quarterly*, Summer.
- DAWSON, V. (2003). The competence of the Australian new graduate speech pathologist. Paper presented at Speech Pathology Australia national conference, Hobart, 4-8 May.
- DREYFUS, H. L. & DREYFUS, S. E. (1996) The relationship of theory and practice in the acquisition of skill. IN BENNER, P. A., TANNER, C. A. & CHESLA, C. A. (Eds.) *Expertise in Nursing Practice: Caring, clinical judgment, and ethics*. New York, Springer Publishing Company.
- DUKE, M. (1996) Clinical evaluation - difficulties experienced by sessional clinical teachers of nursing: A qualitative study. *Journal of Advanced Nursing*, 23, 408-414.
- EMBRETSON, S. E. (1999) Issues in the measurement of cognitive abilities. IN EMBRETSON, S. E. & HERSHBERGER, S. L. (Eds.) *The New Rules of Measurement: What every psychologist and educator should know*. Mahwah, Lawrence Erlbaum Associates, Inc.
- EMBRETSON, S. E. AND REISE, S. P. (2000) *Item Response Theory for Psychologists*. Mahwah, Lawrence Erlbaum Associates, Inc.
- EPSTEIN, R. M. & HUNDERT, E. M. (2002) Defining and assessing professional competence. *Journal of the American Medical Association*, 287, 226-235.
- FAY, C. H. & LATHAM, G. P. (1982) Effects of training and rating scales on rating errors. *Personnel Psychology*, 35, 105-117.
- FERGUSON, A. (2006). Competency-based occupational standards: Influences on Australian speech pathology education. *Folia Phoniatrica et Logopaedica*, 58.23-31.
- FERGUSON, A., LINCOLN, M., MCALLISTER, L. & MCALLISTER, S. (2008) *COMPASSTM Directions: Leading the integration of a competency based assessment tool in speech pathology learning and teaching*. Report prepared for Sydney, Australian Learning & Teaching Council Ltd, an initiative of the Australian Government Department of Education, Employment and Workplace Relations. Accessible from <http://www.altc.edu.au/>
- FRIEDMAN, M. & MENNIN, S. (1991) Rethinking critical issues in performance assessment. *Academic Medicine*, 66, 390 - 395.
- GOMEZ-MEJIA, G. (1988) Evaluating employee performance: Does the appraisal instrument make a difference? *Journal of Organizational Behavior Management*, 9, 155-172.

- GOVAERTS, J. J. B., VAN DER VLEUTEN, C. & SCHUWIRTH, L. W. T. (2002) Optimising the reproducibility of a performance-based test in midwifery education. *Advances in Health Sciences Education*, 7, 133-145.
- GOVAERTS, M. J. B., VAN DER VLEUTEN, C. P. M., SCHUWIRTH, L. W. T. & MUIJTJENS, A. M. M. (2006) Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education*.
- HAGER, P. (2000) Know-how and workplace practical judgement. *Journal of Philosophy of Education* 34(2): 281-296.
- HAGER, P., ATHANASOU, J. & GONCZI, A. (1994) *Assessment Technical Manual*, Canberra, Australian Government Publishing Service.
- HAGE, P. & GILLIS, S. (1995). *Assessment at higher levels. Key aspects of competency-based assessment*, ed. by W.C. Hall, 59-71. Adelaide, SA: NCVER.
- HARRIS, R., GUTHRIE, H., HOBART, B. & LUNDBERG, D. (1995) *Competency-based Education and Training: Between a rock and a whirlpool*, Sydney, Macmillan Education Australia Pty Ltd.
- HAYS, R. B., DAVIES, H. A., BEARD, J. D., CALDON, L. F. M., FARMER, E. A., FINUCANE, P. M., MCCRORIE, P., NEWBLE, D., SCHUWIRTH, L. W. T. & SIBBALD, G. R. (2002) Selecting performance assessment methods for experienced physicians. *Medical Education*, 36, 910-917.
- HIGGS, J. & BITHELL, C. (2001) Professional expertise. IN HIGGS, J. & TITCHEN, A. (Eds.) *Practice Knowledge and Expertise in the Health Professions*. Oxford, Butterworth-Heinemann.
- HUNT, D. D. (1992) Functional and dysfunctional characteristics of the prevailing model of clinical evaluation systems in North American medical schools. *Academic Medicine*, 67, 254-259.
- ILOTT, I. & MURPHY, R. (1997) Feelings and failing in professional training: The assessor's dilemma. *Assessment and Evaluation in Higher Education*, 22, 307 - 316.
- JOHNSON, C. J. AND C. M. SHEWAN (1988) A new perspective in evaluating clinical effectiveness: The UWO clinical grading system. *Journal of Speech and Hearing Disorders* 53, 328-340.
- JONES, A. (2000) The place of judgment in competency-based assessment. *Journal of Vocational Education and Training*, 51, 145-160.
- KEEN, A. J. A., KLEIN, S. & ALEXANDER, D. A. (2003) Assessing the communication skills of doctors in training: Reliability and sources of error. *Advances in Health Sciences Education*, 8, 5-16.
- KINGSTROM, P. O. & BASS, A. R. (1981) A critical analysis of studies comparing Behaviorally Anchored Rating Scales (BARS) and other rating formats. *Personnel Psychology*, 34, 263-289.
- LANDON, B. E., NORMAND, S.-L. T., BLUMENTHAL, D. & DALEY, J. (2003) Physician Clinical Performance Assessment. *Journal of the American Medical Association*, 9, 1183 - 1189.
- LANDY, F. J. & FARR, J. L. (1980) Performance rating. *Psychological Bulletin*, 87, 72-107.

- LEACH, L., NEUTZE, G. & ZEPKE, N. (2001) Assessment and empowerment: Some critical questions. *Assessment & Evaluation in Higher Education*, 26, 293-305.
- LINACRE, J. M. (1994) Sample size and item calibration stability. *Rasch Measurement Transactions [electronic journal]*, 2002.
- LINACRE, J. M. (2002) Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85-106.
- LINCOLN, M., FERGUSON, A., MCALLISTER, L. & MCALLISTER, S. (2008) *Benchmarking clinical learning in speech pathology to support assessment, discipline standards, teaching innovation and student learning*. Australian Learning & Teaching Council Ltd, an initiative of the Australian Government Department of Education, Employment and Workplace Relations. Accessible from <http://www.altc.edu.au/>
- LINN, R. L., BAKER, E. L. & DUNBAR, S. B. (1991) Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20, 15-21.
- MCALLISTER, S. (2005) Competency Based Assessment of Speech Pathology Students' Performance in the Workplace. *Faculty of Health Sciences*. Sydney, The University of Sydney. Accessible from <http://hdl.handle.net/2123/1130>
- MCALLISTER, S. (2008) Introduction to the use of Rasch analysis to assess patient performance. *International Journal of Therapy and Rehabilitation* 15 (11).
- MCALLISTER, S., LINCOLN, M., FERGUSON, A. & MCALLISTER, L. (2006) *COMPASS™: Competency Assessment in Speech Pathology*, Melbourne, Speech Pathology Association of Australia Ltd.
- MCALLISTER, S., LINCOLN, M., FERGUSON, A. & MCALLISTER, L. (2008) Engaging speech pathology students in design and validation of competency based assessment in the workplace. *ATN Assessment Conference*. Adelaide.
- MCALLISTER, S. M., LINCOLN, M., FERGUSON, A. & MCALLISTER, L. (2004) User Evaluation of the Content and Processes of an Assessment of Student Performance in Workplace Settings. *IALP World Congress, 2004, Brisbane, Australia*.
- MCGAGHIE, W. C. (1993) Evaluating competence for professional practice. IN CURRY, L. & WERGIN, J. F. (Eds.) *Educating Professionals*. San Francisco, Jossey-Bass.
- MICHELL, J. (1997) Quantitative science and the definition of *measurement* in psychology. *British Journal of Psychology*, 88, 355-383.
- MOSS, P. (1992) Shifting conceptions of validity in educational measurement: implications for performance assessment. *Review of Educational Research*, 62, 229-258.
- NEWBLE, D., NORMAN, G. & VAN DER VLEUTEN, C. (2000) Assessing clinical reasoning. IN HIGGS, J. & JONES, M. (Eds.) *Clinical Reasoning in the Health Professions*. 2nd ed. Oxford, Butterworth-Heinemann.
- NORCINI, J. J. (2005) Current perspectives in assessment: the assessment of performance at work. *Medical Education*, 39, 880-889.
- NORMAN, G., VAN DER VLEUTEN, C. & DE GRAAFF, E. (1991) Pitfalls in the pursuit of objectivity: Issues of reliability. *Medical Education*, 25, 110-118.

- PETERS, G., FRASER, J., COWIE, F., LOADER, J., RUTTER, S., SCOTT, R. & DAVIE, B. (2001) Competency based assessment in a perioperative nursing graduate diploma. *ACORN Journal*, 15, 11-14.
- RETHANS, J.-J., NORCINI, J., BARON-MALDONADO, M., BLACKMORE, D., JOLLY, B. C., LADUCA, T., LEW, S. R., PAGE, G. G. & SOUTHGATE, L. J. (2002) The relationship between competence and performance: Implications for assessing practice performance. *Medical Education*, 36, 901-909.
- ROBERTSON, S., ROSENTHAL, J. & DAWSON, V. (1997) Using assessment to promote student learning. *Facilitating Learning in Clinical Settings*. Cheltenham, Stanley Thornes Ltd.
- ROSE, M., & BEST, D. (Eds.) (2005) *Transforming practice through clinical education, professional supervision and mentoring*. Edinburgh: Elsevier Churchill Livingstone.
- RCSLT. (2007) *Newly qualified practitioners competency framework: Speech and language therapy competency framework to guide transition to certified RCSLT membership*: Royal College of Speech & Language Therapists (<http://www.rcslt.org/resources/newlyqualifiedpractitioners>, accessed 11 October, 2008).
- SCHUWIRTH, L. W. T. & VAN DER VLEUTEN, C. P. M. (2006) A plea for new psychometric models in educational assessment. *Medical Education*, 40, 296-300.
- SHAVELSON, R. J., GAO, X. & BAXTER, G. (1993) Sampling Variability in Performance Assessments: CSE Technical Report 361. Santa Barbara, National Center for Research on Evaluation, Standards, and Student Testing, University of California.
- SPAA (2001) Competency-Based Occupational Standards for Speech Pathologists (Entry Level). Melbourne, Speech Pathology Association of Australia Ltd.
- SPIKE, N., ALEXANDER, H., ELLIOTT, S., HAZLETT, C., KILMINSTER, S., PRIDEAUX, D. & ROBERTS, T. (2000) In-training assessment - its potential in enhancing clinical teaching. *Medical Education*, 858-861.
- WASS, V., VAN DER VLEUTEN, C., SHATZER, J. & JONES, J. (2001) Assessment of clinical competence. *The Lancet*, 357, 945-949.
- WILKINSON, T. J. & FRAMPTON, C. M. (2003) Assessing performance in final year medical students. Can a postgraduate measure be used in an undergraduate setting? *Medical Education*, 37, 233-240.
- WOEHR, D. J. & HUFFCUTT, A. I. (1994) Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189-205.
- WOLF, A. (1995) *Competence-based assessment*. Philadelphia: Open University Press.
- WOLFE, E. W. & GITOMER, D. H. (2001) The influence of changes in assessment design on the psychometric quality of scores. *Applied Measurement in Education*, 14, 91-107.
- WRIGHT, B. D. (1999) Fundamental measurement for psychology. IN EMBRETSON, S. E. & HERSHBERGER, S. L. (Eds.) *The New Rules of Measurement: What every psychologist and educator should know*. Mahwah, Lawrence Erlbaum Associates, Inc.

ZHU, W. (1996) Should total scores from a rating scale be used directly? *Research Quarterly for Exercise and Sport*, 67, 363-372.