

Research Article

Reliability of the Perceive, Recall, Plan and Perform System of Task Analysis: A criterion-referenced assessment

Melissa T. Nott, Christine Chapparo and Robert Heard

Faculty of Health Sciences, The University of Sydney, Sydney, New South Wales, Australia

Objective: To conduct preliminary examination of the rater and test-reliability of the Perceive, Recall, Plan and Perform (PRPP) System of Task Analysis, an ecological measure designed to assess task-embedded information processing capacity during occupational therapy assessment of confused and agitated adults following traumatic brain injury.

Methods: Occupational therapists observed and scored client performance using the PRPP System of Task Analysis. Correlational analysis and measures of agreement were performed to determine interrater and intrarater reliability. Test procedures were examined for reliability and internal consistency.

Results: Interrater and test reliability considered three factors: therapists, clients and tasks. A moderate level of interrater reliability was achieved between trained therapists (intraclass correlation coefficient (ICC) = 0.60). Test procedures were highly reliable (ICC = 0.88). Across two measurement occasions, therapists showed a tendency towards harder rating on the second test occasion (\bar{d} -4.5%; 95% confidence interval for \bar{d} : -10.67% → 3.17%).

Conclusion: The findings of this study support the use of criterion-referenced tests in the area of occupational performance measurement. Occupational therapists achieved moderate interrater reliability when measuring the performance of adults with brain injury on various activities of daily living. Test procedures were found to be highly

reliable in measuring the occupational performance of adults demonstrating confusion and agitation typical to the stage of post-traumatic amnesia following head injury.

KEY WORDS agitation, assessment, brain injuries, occupational performance, reliability.

Introduction

Measuring change in occupational performance of agitated adults following a traumatic brain injury (TBI) is a complex process. Agitated behaviour is associated with the post-traumatic amnesia (PTA) stage of recovery, and defined as an excess of behaviour characterised by disinhibition, aggression, restlessness or emotional lability (Corrigan & Bogner, 1994), that may include inappropriate vocalising, intolerance of medical management or equipment, and directed or diffuse aggression (Lombard & Zafonte, 2005). The confusion, disorientation, disinhibition and amnesia common to the PTA recovery phase often results in anxiety, agitation, non-compliance and combative behaviour (Slifer *et al.*, 1997) congruous with Level IV of the Rancho Levels of Cognitive Functioning Scale (Hagen, 2001). These cognitive and behavioural features of PTA, including agitated behaviour limit the patient's ability to engage in formal or standardised assessment procedures (Fry & O'Brien, 2002). In addition, many standardised cognitive assessments are deemed invalid during the period of PTA. Assessment of cognition and occupation performance in adults with agitated behaviour during PTA has to date received little investigation.

Occupational therapists require assessment methods that clearly distinguish the role of cognition in the performance of complex human occupations, particularly during the acute phase of recovery from TBI characterised by PTA and concomitant agitated behaviour. Ylvisaker *et al.* (2007) have identified poor outcome measurement and limited reporting of measurement reliability as one of the primary limitations to existing research on behavioural interventions following brain injury. Reliability of observation/measurement is especially critical in behavioural intervention studies because change is rarely measured using standardised tools with established reliability. Of the 65 studies reviewed by Ylvisaker *et al.*,

Melissa T Nott BAppSc (OT) (Hons); PhD Candidate.
Christine Chapparo DipOT (NSW), MA, PhD; Senior Lecturer.
Robert Heard BA (Hons), PhD; Senior Lecturer.

Correspondence: Melissa Nott, J Block, Rm J126, PO Box 170, Lidcombe, NSW 1825, Australia. E-mail: mnot2529@mail.usyd.edu.au

Institution at which work was carried out: Royal Perth Hospital, Shenton Park Campus, 6 Selby Street, Shenton Park, WA 6008, Australia.

Accepted for publication 5 June 2008.

© 2009 The Authors

Journal compilation © 2009 Australian Association of Occupational Therapists

89% used self-developed, context-specific behaviour counts or intensity measures to determine outcome. Only 8% used a standardised measure with documented reliability.

In view of the paucity of outcome measures suitable for use with adults following TBI who demonstrate behavioural and cognitive deficits, the current study is a preliminary investigation into the measurement properties of an occupational performance assessment tool, shown to be sensitive in measuring change over time in adults with severely agitated behaviour during PTA (Nott & Chapparo, 2007).

The Perceive, Recall, Plan and Perform (PRPP) System of Task Analysis is a process-orientated, criterion-referenced assessment that employs task analysis methods to determine problems with cognitive information processing in the context of tasks and activities that are meaningful and relevant to the person (Chapparo & Ranka, 1997a). An underlying assumption of the assessment system is that a person's capacity to process the cognitive demands inherent in everyday tasks can be observed, identified and used to determine the need for occupational therapy intervention (Chapparo & Ranka). The purpose of the assessment is to identify difficulties in specific information processing strategies during task performance and to provide a focus for intervention (Fry & O'Brien, 2002; Nott & Chapparo, 2007). There has been little investigation into the instrument's reliability when used to measure performance of people who are confused and demonstrating agitated behaviour, characteristic of the PTA stage in brain injury recovery.

The PRPP System is an emerging assessment system that measures both task and information processing performance over time and in context, affording therapists a mechanism to measure intervention outcomes at both the level of body structure and function and at the level of activity and participation. This assessment system is particularly relevant for assessing adults following TBI during PTA, as the assessment is observation based. It utilises task performance as the context of the assessment, rather than the use of written or language-based assessments, which can have limited application to adults with TBI during the period of PTA (Fry & O'Brien, 2002). It enables simultaneous assessment of occupational performance and the underlying cognitive performance capacities that may afford or inhibit performance, where occupational performance is defined as carrying out roles, routines and tasks for the purpose of self-maintenance, productivity, leisure and rest in response to demands of the internal and/or external environment (Chapparo & Ranka, 1997a, 1997b). This contrasts with assessments measuring cognitive and/or perceptual skills in isolation from daily occupations; for example the Rivermead Perceptual Assessment Battery (Whiting, Lincoln, Bhavnani & Cockburn, 1985) and the Rivermead Behavioural Memory Test (Wilson, Cockburn, & Baddeley, 2003). It is distinct from other functional evaluations using task observation (for example Arnadottir, 1990; Baum &

Edwards, 1993) in its synthesis of information processing theory and occupational performance. It also contrasts with other ecological assessments such as the Assessment of Motor and Process Skills (AMPS) which is designed to measure activity participation and restriction without specifically evaluating underlying impairments or capacities within the cognitive domain (Boman, Lindstedt, Hemmingsson & Bartfai, 2004).

Development of the PRPP System of Task Analysis

The PRPP System uses a two-stage analysis process. Stage 1 employs a standard *behavioural* task analysis, whereby errors in everyday task performance are identified, generating an overall measure of mastery for specific and relevant occupations. Stage 2 focuses on information processing strategies required for performance by using a *cognitive* task analysis and is the primary focus of this paper. Cognitive task analysis is a family of assessment methods that describe the cognitive processes that underlie performance of tasks and the cognitive strategies used to respond adeptly to complex situations (Schraagen, Chipman & Shalin, 2000).

The conceptual model underlying the PRPP System of Task Analysis was adapted from an information processing model in the field of instructional design used to explain the process of learning tasks in the work place (Romiszowski, 1984). The PRPP model (Fig. 1) is centred around four quadrants that form the inner layer with multidirectional arrows that mirror the multistaged flow of information in theoretical models of information processing. These quadrants categorise cognitive processing strategies used during task performance into four areas: sensory perception (Perceive), memory (Recall), response planning and evaluation (Plan) and performance monitoring (Perform). These are depicted as the inner layer of the PRPP System (Fig. 1).

Each quadrant is broken down into three subquadrants (middle layer Fig. 1) and the underlying information processing strategies that support each of these areas of cognitive function (forming the outer layer Fig. 1). The assessment requires therapists to observe occupational performance, for example, a patient brushing his/her teeth or hair in the bathroom, and to systematically rate the extent to which the patient demonstrates each information processing strategy, as required for that task. A three-point rating scale is used to score the patient's performance of each processing strategy as (3) effective for task performance, (2) questionable or (1) not effective.

The PRPP is a criterion-referenced measure. Criterion-referenced measures assess what clients 'can do and what they know, not how they compare to others' (Anastasi, 1988, p. 102). In Stage 2 of the PRPP assessment, the processing strategies that form the outer ring of the model are specified as the criteria, which must be demonstrated relative to a predetermined performance level. The

required between standardised assessment procedures and realistic implementation in practice settings (Sudsawad, 2005). In criterion-referenced measures, such as the PRPP, this is achieved through using real-life contextual performance as the outcome of interest to increase the meaningfulness and relevance of findings for occupational therapy practice. The goal of the PRPP System, as a criterion-referenced measure, is to obtain a description of the specific information processing skills each patient can demonstrate, with no attempt to eliminate easy items or alter their difficulty.

While criterion-referenced measures may be ecologically valid, they present difficulties when attempting to determine the reliability of measuring human performance. Highly reliable assessments that focus on 'can/cannot' or 'mastery/non-mastery' interpretation often reduce the assessment to the level of trivial sets of skills that, while highly reliable, are of little use in interpreting overall performance (Griffin, 1995). Norm-referenced measures contain highly reliable, stable performance outcomes against which client performance can be judged; however, these are largely removed from everyday function. Moreover, most everyday tasks have not one single competency, but many. Competence for everyday function is not stable. In real world contexts, it depends on how well people can adapt to changes in the performance context, their capacities, and the complexity of task demands (Griffin). Evaluation of function involves recognition that task performance does not have one outcome, approach or solution only (Glaser, 1981; Griffin), and that assessor knowledge and background have an impact on the reliability of scores derived among assessors such as occupational therapists, who are assumed to have the same understanding of the performance criterion, the disability or illness inhibitors to performance. In these instances, the importance of well-defined 'rules' for judging performance, termed scoring 'rubrics' is paramount (Eisner, 1993).

A growing body of literature documents the use of criterion referencing in rehabilitation (see for example Rockwood, Joyce & Stolee, 1997) and education (see for example Carlson, MacDonald, Gorely, Hanrahan & Burgess-Limerick, 2000). Although the use of the PRPP System of Task Analysis is expanding in clinical practice and in research, the measurement properties of this criterion-referenced measure have not been widely published. A number of small, unpublished studies have been conducted across various client groups that support its use. For example, several studies report interrater and test-retest reliability ranging from 0.64 to 0.99 (Lohri, 2005; Munkhetvit, 2005). This study further investigated inter- and intrarater reliability of the PRPP System of Task Analysis for measuring information processing in adults with brain injury, in PTA, demonstrating agitated behaviour during self-care task performance, and the reliability of PRPP assessment procedures.

Methods

This study was conducted with the approval of local human ethics committees. Informed consent was obtained from legal guardians for each patient videotaped, and from participant raters (occupational therapists).

Participants

Nine occupational therapists who had previously been trained to use the PRPP System of Task Analysis participated in the study. Recruitment occurred on a voluntary basis in response to advertising in an Australian national occupational therapy newsletter and via internet list serves in two states. The mean age of therapists was 31 years (standard deviation (SD) = 8; range 23–50). On average, therapists had participated in PRPP training 3.4 years prior to the study (SD = 1.8; range 1–6). One therapist reported using the PRPP System daily, three reported monthly use, and five reported occasional use. Half the sample ($n = 4$) worked in the practice area of adult TBI rehabilitation, one third ($n = 3$) worked in adult acute neurology, with the remaining sample working in paediatrics ($n = 1$), and in a community-based/case management role ($n = 1$). All participants had undergone training to use the PRPP System of Task Analysis involving a five-day training workshop, and had experience working with adults with brain injury.

Patients who met the following selection criteria were considered for the study. Adults older than 18 years, admitted for neurosurgical rehabilitation following a brain injury and demonstrating agitation and confusion, were identified by the treating occupational therapist. Clinical assessment of agitation was confirmed using the Rancho Levels of Cognitive Functioning Scale (Hagen, 2001). Patients demonstrating behaviours consistent with Rancho Level IV were recruited to the study.

Procedures

Five patients (four males, one female), aged 19–58 years, who met the above selection criteria were recruited to the study. Patients were videotaped performing self-care tasks during occupational therapy sessions. Tasks were chosen by the treating occupational therapist in accordance with rehabilitation goals at the time of filming and included eating, drinking, face washing, teeth brushing, hair brushing and upper-body dressing. The treating occupational therapist and researcher jointly conducted each 45-min therapy session. During filming, a least-to-most hierarchy of prompting and assistance was provided to ensure completion of each task with the least assistance from staff. Videotaped footage of the patients' performance was copied to digital form for scoring. Research packages containing digital footage and PRPP scoring sheets were distributed by mail to all participants, who independently scored and returned these within a 6 to 8-week period. A second DVD containing one client, performing one task, was sent two weeks later to all

participants for the purposes of measuring intrarater agreement. The use of videotapes of patients performing activities has been shown to be effective in allowing multiple raters to observe and score the exact same performance (Portney & Watkins, 2000).

Data analysis

Interrater and test reliability

Interrater reliability was concerned with the variation between the raters who measured the same group of people, while test reliability was focussed on the PRPP System itself and estimated how reliably the PRPP System measured the population from which the sample of patients came. A three-way intraclass correlation coefficient (ICC) (Wong & McGraw, 1999) was used to estimate interrater and test reliability. The three factors in this type of analysis are traditionally called persons (p), items (i) and judges (j). In the current study, these factors were termed patients, tasks and raters, respectively.

Nine raters each assessed the same five patients on four tasks, totalling 180 ratings. The PRPP total percentage scores (summed total of all descriptors) and individual quadrant scores are used in these correlational analyses. Each patient was assessed on the four tasks most appropriate to him or her, meaning that tasks were not the same for every patient. In statistical terms, the factor of task was nested within the factor of patient, written as $(i:p) \times j$ in Wong and McGraw's terminology.

Two ICCs were calculated for each of the four PRPP quadrant scores, and the PRPP total score. The first ICC was focussed on the raters, and was structured to assess the reliability of an average rater in the population from which the raters came. The second ICC was focussed on the PRPP test procedures, and was structured to assess the reliability of the PRPP assessment system as applied to the population from which the sample of patients came. This second ICC is analogous to the calculation of Cronbach's alpha, for measuring the internal consistency of a test. Three-way ICCs and their confidence intervals (CI) were calculated using a spreadsheet provided by McGraw based on the mathematical equations presented in Wong & McGraw (1999). ICCs were calculated based on data produced from a type IV sum of squares analysis of variance (ANOVA), where patients were nested with tasks, and crossed by all raters. In accordance with other measures of agreement, ICC = 0.41–0.60 is moderate agreement, and 0.61–0.80 is considered substantial agreement between raters or testing occasions (Landis & Koch, 1977). Measures of test reliability are interpreted as for Cronbach's alpha requiring at least 0.70 or higher for an adequate scale, and a cut-off of 0.80 for a good scale.

Intrarater agreement

Intrarater agreement refers to the stability of measures recorded by one rater across two or more trials, and was measured for each individual rater following a 2-week interval between scoring occasions. Analysis of intrarater

agreement and disagreement was conducted using the Bland and Altman method to assess therapists' ratings on repeated scoring occasions (Bland & Altman, 1986). Calculations included the mean of scores at time 1 and time 2, against the difference between scores at time 1 and time 2, the 95% confidence intervals for this difference, and the SD of the difference (SD_{diff}). A diagram was plotted to illustrate the distribution of results. In this study, mean total PRPP measures from time 1 and time 2 were plotted on the x -axis. The y -axis displays the difference between total PRPP measures at scoring time 1 and scoring time 2 (time 2–time 1). If ratings were completely reliable, the difference between therapist's ratings at time 1 and time 2 would be zero (0). If a therapist's time 2 rating was higher than his/her time 1 rating, the difference would be a positive value. Conversely, if a therapist's time 2 rating was lower than his/her time 1 rating, the difference would be a negative value. If therapists generally rated higher or lower at time 2, the average difference would be significantly different to 0, a situation referred to as 'bias'. The presence of bias was tested by calculating the SD and 95% CI of the difference between ratings at scoring time 1 and scoring time 2. We expect 95% of the difference to be less than 2 SDs from the mean (Bland & Altman). The potential for rater bias increases when using observational methods, and is best managed by developing objective scoring criteria and by rigorously training the raters (Portney & Watkins, 2000).

Results

Interrater reliability

The ICC coefficient for interrater reliability based on a three-way model was moderate, with a reliability estimate of 0.60 (see Table 1 for CIs). Each quadrant of the PRPP System of Task Analysis was then measured separately to determine the relative contribution of each quadrant to the reliability estimate. The quadrant-by-quadrant analysis indicated interrater reliability ranging from 0.51 to 0.59.

Test reliability

The ICC coefficient for test reliability based on a three-way model was very high, with a reliability estimate of 0.88, ranging from 0.83 to 0.88 for each quadrant (see Table 1 for confidence intervals) meeting the above-identified cut-off (0.80) for achieving a 'good scale'.

Intrarater agreement

The Bland and Altman test results are presented in Table 2. The mean percentage difference from time 1 to time 2 was -4.52 , indicating a tendency towards lower ratings at time 2. A bias towards negative differences is evident in Figure 2, with the distribution of data points primarily below the zero reference line, the point at which the difference between scores from time 1 to time 2 is zero

TABLE 1: Intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) for interrater reliability and test reliability

	ICC raters	95% CI	ICC test	95% CI
PRPP total	0.60	0.25 → 0.93	0.88	0.64 → 0.98
Perceive quadrant	0.59	0.26 → 0.93	0.88	0.64 → 0.99
Recall quadrant	0.59	0.23 → 0.93	0.86	0.59 → 0.98
Plan quadrant	0.51	0.18 → 0.91	0.83	0.51 → 0.98
Perform quadrant	0.53	0.19 → 0.91	0.88	0.65 → 0.99

(indicating perfect agreement). Six points lie below zero indicating a negative value for the difference between scoring time 1 and scoring time 2. A trend towards negative differences when the mean average is higher is also apparent from the plot. The 95% CI for mean difference was calculated to $-10.67 \rightarrow 3.17$, falling well within the 2 SD range of the mean difference (-4.52 ± 18.00), and crossed zero (the expected difference between time 1 and time 2 if exact agreement was achieved).

Discussion

This study measured intrarater agreement, interrater reliability and test reliability of the PRPP System of Task Analysis, using data from the PRPP Stage 2 analysis. Criterion-referenced assessments have previously been criticised for their lack of reliability between raters and inherent subjectivity of criteria interpretation. In this study, the PRPP System of Task Analysis was seen to have moderate interrater reliability when used to measure agreement between trained therapists and good test reliability.

Occupational therapists have limited choice in assessment tools appropriate for use with adults in PTA following brain injury. The added complication of agitated behaviour during task performance prevents use of pen-and-paper-based screening tools or full assessment batteries. Beyond daily monitoring of PTA status, limited cognitive assessment occurs in the current clinical context of acute

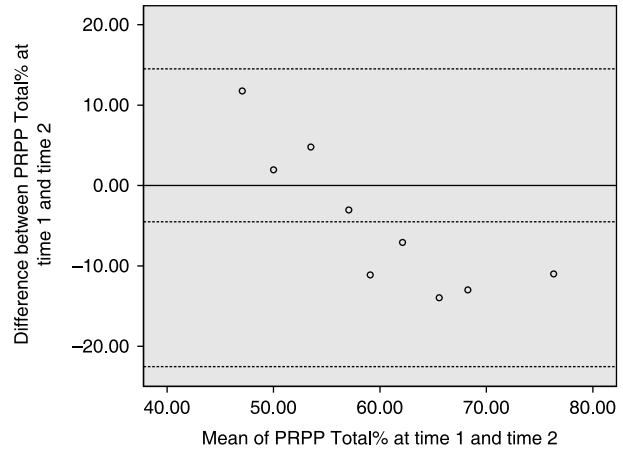


FIGURE 2: Bland and Altman plot of differences between time 1 and time 2 with mean Perceive, Recall, Plan and Perform (PRPP) total percentage (Total%) scores with zero reference line (solid), mean difference and ± 2 standard deviation (SD) lines (dotted).

brain injury rehabilitation, prior to PTA emergence. Functional observation using structured methods for recording and measuring behaviour are effective in measuring change during PTA (Weir, Doig, Fleming, Wiemers & Zemljic, 2006). Tools that combine structured observation and measurement of task-specific information processing with an overall measure of functional performance, such as The PRPP System of Task Analysis, offer a unique assessment system to occupational therapists working in this area of clinical practice. The PRPP System has been shown to be sensitive to change during this stage of recovery (Nott & Chapparo, 2007) and instrumental in guiding intervention (Fry & O'Brien, 2002). The criterion-referenced procedures that underpin scoring and administration of the assessment have been shown to offer good reliability across raters and within the assessment procedures themselves.

Criterion-referenced assessments that rely on professional judgement when scoring across different tasks and different clients require good reliability to allay concerns about the transparency and consistency of scoring procedures (Dunn, Morgan, Reilly & Parry, 2004). In this study, the reliability of test procedures was measured concurrently with rater reliability and found to demonstrate

TABLE 2: Bland and Altman tests for intrarater agreement between scoring occasions

	Mean PRPP scores			Bland and Altman		
	Time 1	Time 2	\bar{d}	SD_{diff}	SE of \bar{d}	95% CI for \bar{d}
Intrarater agreement	62.15%	57.63%	-4.52	9.00	3.00	-10.67 → 3.17

CI, confidence interval; PRPP, Perceive, Recall, Plan and Perform; SD, standard deviation; SE, standard error.

good reliability across therapists, tasks and patients. While raters should possess an understanding of the relevance of criteria to both the task being assessed and the person who is performing the task, rigorous scoring rubrics are essential in achieving high levels of test reliability. Therapists undergo intensive postgraduate training focussed on applying the scoring rubric to a range of behaviours, both during the assessment workshop and postworkshop in the clinical setting.

Stability in measurement is important when evaluating patients over time in order to monitor change. Therapists' observations and scoring procedures (intrarater agreement) were measured over a 2-week period, identifying that therapists tended to give slightly lower scores at time 2. Factors leading to this require further investigation; however, some suggestions are offered. The reallocation of a three-rating to a two-rating on one strategy within a subquadrant/quadrant was commonly associated with several lowered scores in the same subquadrant/quadrant. This tendency to allocate same scores to items within the same subquadrant or quadrant may have contributed to a type of 'consistency effect' or 'consistency motif' (Podsakoff, MacKenzie, Lee & Podsakoff, 2003) at time 2. Alternatively, the research study procedure may have contributed to an unexpected source of error during scoring at time 2. On the first scoring occasion, therapists viewed and scored five patients. On the second testing occasion, therapists were required to rescore one patient only. This difference in measurement procedure may have impacted on therapists' rating of this patient, who they may have perceived as performing well in comparison to the other four patients, but more impaired when viewed alone. This may indicate that therapists compared performance *between* patients on the first measurement occasion which was an unexpected source of error, and not the intended purpose of a criterion-referenced measure (Anastasi, 1988). If this did in fact occur, it would have reduced both inter and intrarater reliability.

Limitations

In order to achieve this research design with a three-way correlational model, all therapists are required to observe and rate all clients. In clinical settings, this is only feasible via scoring of video footage. Participants reported that scoring via video reduced the opportunities for therapists to engage directly with the client and created an artificial 'distance' between the therapist and patient that would not usually be present during clinical contexts.

Second, while the small sample size of therapists and clients is a limitation of this study, the standardised procedures and initial examination of reliability across nine therapists is a large step forward in comparison to the self-developed outcome measures typically used to demonstrate change in adults with cognitive and behavioural difficulties following TBI (Ylvisaker *et al.*, 2007). In addition, the findings contribute to the published information available on the PRPP System of Task Analysis,

which is in itself, an emerging assessment method. Generalisation of findings may be limited to therapists with PRPP training, working with a similar client group of adults with brain injury.

Conclusion

The PRPP System of Task Analysis is a two-stage criterion-referenced assessment that measures both task performance skill and cognitive information processing capacity over time and in context. Occupational therapists have limited choice in assessment tools appropriate for use with adults in PTA following brain injury. The PRPP System of Task Analysis offers a unique assessment system to fill this gap for therapists working in this area of clinical practice. The criterion-referenced procedures that underpin scoring and administration of the assessment have been shown to offer good reliability across raters and within the assessment procedures themselves.

References

- Anastasi, A. (1988). *Psychological testing*. New York: MacMillan Publishing Co.
- Arnadottir, G. (1990). *Brain and behavior: Assessing cortical dysfunction through activity of daily living*. St Louis, MO: Mosby.
- Baum, C. & Edwards, D. F. (1993). Cognitive performance in senile dementia of the Alzheimer's type: The kitchen task assessment. *American Journal of Occupational Therapy*, 47, 431–436.
- Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327, 307–310.
- Boman, I., Lindstedt, M., Hemmingsson, H. & Bartfai, A. (2004). Cognitive training in home environment. *Brain Injury*, 18, 985–995.
- Carlson, T., MacDonald, D., Gorely, T., Hanrahan, S. & Burgess-Limerick, R. (2000). Implementing criterion-referenced assessment within a multi-disciplinary university department. *Higher Education Research and Development*, 19, 103–116.
- Chapparo, C. & Ranka, J. (1997a). *The Perceive, Recall, Plan and Perform System of Task Analysis*. In: C. Chapparo & J. Ranka (Eds.). *Occupational Performance Model (Australia), Monograph 1* (pp. 189–198). Sydney, NSW: Total Print Control.
- Chapparo, C. & Ranka, J. (1997b). *Occupational Performance Model (Australia), Monograph 1*. Sydney, NSW: Total Print Control.
- Corrigan, J. D. & Bogner, J. A. (1994). Factor structure of the agitated behavior scale. *Journal of Clinical and Experimental Neuropsychology*, 16, 386–392.
- Dunn, L., Morgan, C., Reilly, M. & Parry, S. (2004). *The student assessment handbook. New directions in traditional and on-line assessment*. London: Taylor & Francis.
- Eisner, E. W. (1993). Reshaping assessment in education: Some criteria in search of practice. *Journal of Curriculum Studies*, 25, 219–233.

- Fry, K. & O'Brien, L. (2002). Using the Perceive, Recall, Plan and Perform system to assess cognitive deficits in adults with traumatic brain injury: A case study. *Australian Occupational Therapy Journal*, 49, 182–187.
- Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 36, 923–936.
- Griffin, P. (1995). Competency assessment: Avoiding the pitfalls from the past. *Australian New Zealand Journal of Vocational Education Research*, 3, 34–59.
- Hagen, C. (2001). The expert's corner. Proper use of the Rancho Levels of Cognitive Functioning. *Re-Learning Times*, 8, 1.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lohri, J. (2005). Information and sensory processing abilities of young children with autism spectrum disorder. Unpublished Thesis. Sydney, NSW: The University of Sydney.
- Lombard, L. A. & Zafonte, R. D. (2005). Agitation after traumatic brain injury: Considerations and treatment options. *American Journal of Physical Medicine and Rehabilitation*, 84, 797–812.
- Munkhetvit, P. (2005). Cognitive assessment in patients with acquired brain injury in Thailand. Unpublished PhD Thesis. Sydney, NSW: The University of Sydney.
- Nott, M. T. & Chapparo, C. (2007). Measuring information processing in a client with extreme agitation following traumatic brain injury using the Perceive, Recall, Plan and Perform System of Task Analysis. *Australian Occupational Therapy Journal*, 55, 188–198.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended practices. *Journal of Applied Psychology*, 88, 879–903.
- Portney, L. G. & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Rockwood, K., Joyce, B. & Stolee, P. (1997). Use of goal attainment scaling in measuring clinically important change in cognitive rehabilitation patients. *Journal of Clinical Epidemiology*, 50, 581–588.
- Romiszowski, A. (1984). *Designing instructional systems*. London: Hogan Page.
- Schraagen, J. M., Chipman, S. F. & Shalin, V. (2000). *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Slifer, K. J., Tucker, C. L., Gerson, A. C., Sevier, R. C., Kane, A. C., Amari, A. et al. (1997). Antecedent management and compliance training improve adolescents' participation in early brain injury rehabilitation. *Brain Injury*, 11, 877–889.
- Sudsawad, P. (2005). Concepts in clinical scholarship. A conceptual framework to increase usability of outcome research for evidence-based practice. *American Journal of Occupational Therapy*, 59, 351–355.
- Weir, N., Doig, E. J., Fleming, J. M., Wiemers, A. & Zemljic, C. (2006). Objective and behavioural assessment of the emergence from post-traumatic amnesia (PTA). *Brain Injury*, 20, 927–935.
- Whiting, S., Lincoln, N. B., Bhavnani, G. & Cockburn, J. (1985). *The Rivermead perceptual assessment battery*. Windsor UK: NFER – Nelson.
- Wilson, B. A., Cockburn, J. & Baddeley, A. (2003). *Rivermead Behavioural Memory Test II*. Bury St Edmunds, UK: Thames Valley Test Company.
- Wong, S. P. & McGraw, K. O. (1999). Confidence intervals and *F* tests for intraclass correlations based on three-way random effects models. *Educational and Psychological Measurement*, 59, 270–288.
- Ylvisaker, M., Turkstra, L., Coehlo, C., Yorkston, K., Kennedy, M., Sohlberg, M. M. et al. (2007). Behavioural interventions for children and adults with behaviour disorders after TBI: A systematic review of the evidence. *Brain Injury*, 21, 769–805.