Feature Article

Reliability of the Australian Therapy Outcome Measures for Occupational Therapy Self-care scale

Fiona Scott,1 Carolyn A. Unsworth,1 Janet Fricke1 and Nicholas Taylor2

1School of Occupational Therapy, and 2School of Physiotherapy, La Trobe University, Melbourne, Victoria, Australia

Background and Aims: With the shift of health care towards evidence-based practice, demonstrating outcomes of intervention has become an important component of occupational therapy practice. Outcome measures and assessment procedures with sound psychometric properties are necessary to demonstrate such therapy outcomes. This study investigated the interrater and intrarater reliability of one of the 12 scales from a newly developed outcome measure, the Australian Therapy Outcome Measures for Occupational Therapy (AusTOMs-OT).

Method and Results: Seven occupational therapists rated 15 written case studies on two occasions on the four domains (impairment, activity limitation, participation restriction and distress/well-being) of the AusTOMs-OT self-care scale. The results showed that the AusTOMs-OT self-care scale had moderate to high interrater reliability with intraclass correlation coefficients (ICCs) of over 0.79 for the three domains: activity limitation, participation restriction and distress/well-being, and over 0.70 for impairment. Intrarater reliability was also reported to be moderate to high, with ICCs of 0.88 for activity limitation, 0.81 for participation restriction, 0.94 for distress/well-being, and 0.74 for impairment.

Conclusion: The findings of this study support the reliability of the AusTOMs-OT self-care scale and suggest that it can be used to evaluate self-care intervention outcomes, thus contributing to the evidence-based practice of occupational therapy.

KEY WORDS assessment, outcome measure.

Introduction

Occupational therapy intervention aims to restore function, facilitate involvement in activities of choice and participation in life roles, and enhance the well-being of clients (Law, Polatajko, Baptiste & Townsend, 2002). Occupational therapists must provide evidence to show clients, the profession and employers that therapy can achieve these outcomes (Law & Baum, 1998). Outcomes of therapy need to be measured with tools that are reliable, valid and sensitive and that are capable of demonstrating client change in the functional areas targeted during occupational therapy intervention (Donnelly & Carswell, 2002; Dunn, 2001; Hayes, 2000; Kaplan, 1996; Keith, 1995; Ottenbacher, 1997; Rogers & Holm, 1994; Unsworth, 2000). Outcomes measured with such tools can contribute to evidence-based practice. The aim of this study was to investigate the interrater and intrarater reliability of the Australian Therapy Outcome Measures for Occupational Therapy (AusTOMs-OT) self-care scale. The AusTOMs-OT measures a person’s level of impairment, activity limitation, participation restriction and distress/well-being in order to provide a holistic view of their status.

Outcome measures

An outcome measure is a tool that can record a client’s status in relation to specified attributes at two (or more) points in time, usually pre- and post-intervention. Hence, outcome measures can record change over time. Outcome measurements can be used clinically or in research to demonstrate the effectiveness of intervention and thus assure the quality of therapy offered, all of which contribute to evidence-based practice (Dunn, 2001; Fricke, 1993; Unsworth, 2000). The International Classification of Functioning Disability and Health (ICF) (WHO, 2002) provides a useful framework and a global language for describing an individual’s health status, and can be used as a guide when selecting and using outcome measures (Unsworth, 2000). The ICF recognises that disease, along with personal and
environmental factors, all contributes to a person’s experience of impairment (to body structure or function), activity (the execution of a task or action) and participation (involvement in a life situation). The ICF provides a suitable framework for measuring outcomes for all disciplines, however given their training, occupational therapists are well placed to measure outcomes in relation to activity and participation.

A variety of tools can be used to measure health outcomes including the Functional Independence Measure (FIM™) (Uniform Data System for Medical Rehabilitation, 1999), London Handicap Scale (LHS) (Harwood, 1995), Nottingham Health Profile (NHP) (Hunt, McKewen & McKenna, 1986), Sickness Impact Profile (SIP) (Bergner, Bobbitt, Carter & Gilson, 1981), Short Form Health Survey (SF-36) (Ware & Sherbourne, 1992), Therapy Outcome Measure (TOM) (Enderby & John, 1997; Enderby, John & Petherman, 1998), Older Americans Resources and Services Multidimensional Functional Assessment Questionnaire (OARS-MFAQ) (Fillenbaum, 1988), and Personal Care: Participation Outcome Measure (PC-PART) (Darzins, 2004; Vertesi, Darzins, Lowe, McEvoy & Edwards, 2000). While these tools offer a variety of uses, many are of limited value for use as a comprehensive outcome measure. Several tools contain items that are irrelevant to occupational therapy practice (many were not developed for or by occupational therapists) (LHS, NHP, OARS, SIP, SF-36, TOM).

2. The tools do not measure the full scope of occupational therapy intervention (all tools reviewed) (in relation to points 1 and 2, such assessments may be of limited value in measuring change relating to the full scope of occupational therapy intervention).

3. With the exception of TOM (Enderby & John, 1997; Enderby et al., 1998), the tools do not measure all the domains of a client’s function (impairment, activity limitation, participation restriction and distress/well-being).

4. Several tools are self-report or interview, and while self-report could be combined with observation for the participation and distress/well-being domains of the AusTOMs, research has indicated that clinician observation of the client’s function is the most objective method of data collection (Edwards, 1990; Skruppy, 1993) (LHS, NHP, OARS, SIP, SF-36). Objective measurement maintains consistency to provide more reliable outcomes.

5. Length of time to complete (PC-PART, OARS).

6. Complex scoring procedures or interpretation (NHP, SF-36).

Ideally, a new outcome measure is required for occupational therapy that is quick and easy to use and score, based on direct observation and reflects knowledge and skills within the scope of occupational therapy practice. The Australian Therapy Outcome Measures (AusTOMs) (Perry et al., 2004) was developed by and for occupational therapists, physiotherapists and speech pathologists as a result of this need.

Development of AusTOMs

The AusTOMs was developed using the TOM (Enderby et al., 1998) and the ICF (WHO, 2002). Separate scales were developed for the three professions of occupational therapy (AusTOMs-OT), physiotherapy (AusTOMs-PT) and speech pathology (AusTOMs-SP). The scale headings and content were based on ICF concepts, and the use of four domains (impairment, activity limitation, participation restriction and distress/well-being) and an 11-point scale (six defined points from 0 to 5, and five half-points) were adopted from the TOM. The use of a common ICF framework but separate scales for the occupational therapist, physiotherapist and speech pathologist enables a common assessment language across interdisciplinary teams. The 12 AusTOMs-OT scales are: (i) learning and applying knowledge, (ii) functional walking and mobility, (iii) upper limb use, (iv) carrying out daily life tasks and routines, (v) transfers, (vi) using transport, (vii) self-care, (viii) domestic life — home, (ix) domestic life — managing resources, (x) interpersonal interactions and relationships, (xi) work, employment and education, and (xii) community life, recreation, leisure and play. The development of the AusTOMs has been fully described for all three professions in Perry et al. (2004) and for occupational therapy in Unsworth (2005). The scales and manual (Unsworth & Duncombe, 2004) may be ordered through www.latrobe.edu.au/automs.

The AusTOMs-OT scales were designed to be incorporated into the usual routine of occupational therapy (Unsworth & Duncombe, 2004). The therapist gathers information in the usual fashion and therapy goals are set. Prior to commencing therapy the client is scored on all four domains on the AusTOMs-OT scale(s) that are relevant to current therapy goals. Hence, clients are not usually scored on all 12 scales. The intervention program is implemented and therapy goals are continually re-evaluated using the therapist’s usual methods. Once a goal is achieved or once intervention ceases, the client is again scored on the relevant scale(s) (Unsworth, 2005).

Psychometric properties of the AusTOMs-OT

For therapists to confidently adopt an outcome measure, the tool needs to be standardised, including published data for comparison of the client’s score, details
on known reliability and validity, and operational definitions (de Clive-Lowe, 1996; Donnelly & Carswell, 2002). The AusTOMs-OT manual provides these details (Unsworth & Duncombe, 2004). Clear operational definitions are vital for achieving high reliability (Fricke, 1993), particularly with highly sensitive tools. For the four domains of the 12 occupational therapy scales, Wilcoxon signed ranks tests ranged from \( Z = -2.280 \) to \( Z = -12.186 \) and were all statistically significant \( (P < 0.05) \), supporting the sensitivity of all 12 scales to change in client status over time (Unsworth, 2005).

A construct (concurrent) validity study was undertaken by comparing client data using the AusTOMs with data collected using the EuroQuol-5D (EQ-5D) (Brooks, 1996). Moderate to strong statistically significant Spearman Rho correlations between the AusTOMs-OT and EQ-5D were found across all four domains ranging from 0.612 to 0.748 and these data provide preliminary evidence that the AusTOMs and EQ-5D are both measuring global health outcomes (Unsworth et al., 2004). Future validity studies could investigate the ability of AusTOMs-OT to predict client discharge data from admission status, and to determine the capacity of the tool to discriminate between clients with differing severity levels of impairment, activity limitation, participation restriction and distress/well-being.

Reliability refers to how well a tool shows true variation between individuals (Nunnally, 1970); the consistency of a tool when the test is repeated (Kaplan, 1996; Ottenbacher, 1997; Portney & Watkins, 2000); and the amount of error that exists in a tool (Streiner & Norman, 1995). While the reliability and validity of an outcome measure are equally important ‘high reliability is a necessary but not sufficient condition for high validity’ (Nunnally, 1970; p. 107).

There are various types of reliability, however, interrater and intrarater reliability are considered the most important for outcome measures (Ottenbacher, 1997). Interrater reliability is important because raters need to be certain that a change in score is due to the client’s function and not error between the raters. Data gathered from multiple raters with a consistent tool can be pooled and thus larger quantities of data are collected, which can then be used in research or for evaluation of the efficiency and effectiveness of intervention. A measure with good intrarater reliability allows an occupational therapist to accurately measure change in the client over time (Ottenbacher, 1997) and the results can be used for accountability purposes and outcomes research. This contributes to evidence-based practice because the effectiveness of occupational therapy intervention can be demonstrated. To use results in research it is important to understand how much change indicates real change and how to generalise results between therapists and over time.

Preliminary research was conducted into the interrater and intrarater reliability of the 12 AusTOMs-OT scales (Morris et al., 2005). The percentage of agreement over time and between clinicians was calculated and for most scales it was found that for intrarater and test–retest reliability there was 60–100% agreement. Although these results provide preliminary evidence of reliability, the data collected did not allow for more robust data analyses (such as the use of ICCs). To strengthen evidence of reliability it is also important that results are replicated. Hence, further studies such as the one reported in this paper for the self-care scale are required for all AusTOMs-OT scales.

This research examined the reliability of the AusTOMs-OT self-care scale. The self-care scale was selected for study as this was the most common scale rated by occupational therapists in the initial data collection with 466 clients (Unsworth, 2005). The specific aims of the study were to investigate the:

- interrater reliability of the AusTOMs-OT self-care scale,
- intrarater reliability of the AusTOMs-OT self-care scale, and
- retest reliability in the units of measurement to determine what change on the AusTOMs-OT self-care scale is required to be regarded as a ‘real’/clinical change.

**Methods**

**Participants**

Participants were seven occupational therapists at Donvale Rehabilitation Hospital, Victoria Australia, which comprises mostly older adult clientele post-cerebrovascular accident or joint replacement. These occupational therapists were a convenience sample who all consented to participate. Of the seven participants, one was male (14%) and six were female (86%). Participants had a mean age of 32 years (SD 8, range 22–44) and a mean of 7 years experience (SD 6, range 1–17) practising as an occupational therapist. All seven participants were qualified with a Bachelor of Occupational Therapy, two held masters degrees in occupational therapy and one was currently completing a masters degree. To ensure that anonymity was maintained, these demographics were not linked with the participants' ratings.

**Instruments**

*AusTOMs-OT Self-care scale*

The participants rated 15 written case study clients over the four domains of the AusTOMs-OT self-care scale. While the development and scoring of the AusTOMs-OT are described briefly in the introduction,
the reader is referred to Unsworth and Duncombe (2004), Unsworth (2005), and Perry et al. (2004) for full descriptions. The self-care scale is attached as Appendix 1.

Development of client case studies

The 15 written case studies were developed by the first two authors to cover a wide range of client aetiologies and ages. The selection of the medical condition for each of the case studies was based on information from the first AusTOMs-OT data collection of 466 patients at 10 acute, subacute and community facilities across metropolitan Melbourne and one regional facility (Unsworth, 2005). The actual conditions used for the case studies were then randomly selected from the AusTOMs report disorder codes list by the primary researcher. This process was used to ensure that the 15 case studies developed were representative of client aetiologies seen by occupational therapists.

To determine the age of each case study client, five age categories were created (child, adolescent, adult, older adult and old old). The researchers allocated two child cases, one adolescent case, four adult cases, six older adult cases and two old-old cases, to reflect the approximate distribution of client cases seen by clinicians from the initial AusTOMs-OT data collection (Unsworth, 2005). Information on the age, age category, aetiology and aetiology code of the case studies is summarised in Table 1.

Table 1: Summary of case studies

<table>
<thead>
<tr>
<th>Case no. and name</th>
<th>Age (years)</th>
<th>Age category</th>
<th>Aetiology</th>
<th>Aetiology code from AusTOMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 — Bill</td>
<td>86</td>
<td>Old old</td>
<td>Chronic obstructive airways disease</td>
<td>Respiratory disease</td>
</tr>
<tr>
<td>2 — Judy</td>
<td>53</td>
<td>Adult</td>
<td>Osteoarthritis</td>
<td>Musculoskeletal</td>
</tr>
<tr>
<td>3 — Edith</td>
<td>70</td>
<td>Older</td>
<td>Total hip replacement</td>
<td>Orthopaedic</td>
</tr>
<tr>
<td>4 — Jason</td>
<td>17</td>
<td>Adolescent</td>
<td>Spinal cord injury</td>
<td>Orthopaedic</td>
</tr>
<tr>
<td>5 — Dorothy</td>
<td>58</td>
<td>Older</td>
<td>Foot amputation</td>
<td>Amputation</td>
</tr>
<tr>
<td>6 — Claire</td>
<td>37</td>
<td>Adult</td>
<td>Schizophrenia</td>
<td>Mental illness</td>
</tr>
<tr>
<td>7 — Alfred</td>
<td>67</td>
<td>Older</td>
<td>Depression</td>
<td>Mental illness</td>
</tr>
<tr>
<td>8 — James</td>
<td>44</td>
<td>Adult</td>
<td>Acute myocardial infarction</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>9 — Kate</td>
<td>72</td>
<td>Older</td>
<td>Cerebrovascular accident</td>
<td>Acquired neurological</td>
</tr>
<tr>
<td>10 — David</td>
<td>6</td>
<td>Child</td>
<td>Acquired brain injury</td>
<td>Acquired neurological</td>
</tr>
<tr>
<td>11 — Emily</td>
<td>84</td>
<td>Old old</td>
<td>Dementia</td>
<td>Acquired neurological</td>
</tr>
<tr>
<td>12 — Glenda</td>
<td>60</td>
<td>Older</td>
<td>Parkinson’s disease</td>
<td>Progressive neurological disease</td>
</tr>
<tr>
<td>13 — Joe</td>
<td>30</td>
<td>Adult</td>
<td>Burn to upper limb</td>
<td>Burns/Plastics</td>
</tr>
<tr>
<td>14 — Macy</td>
<td>3</td>
<td>Child</td>
<td>Cerebral palsy</td>
<td>Congenital neurological</td>
</tr>
<tr>
<td>15 — Marco</td>
<td>63</td>
<td>Older</td>
<td>Chronic kidney disease</td>
<td>Other</td>
</tr>
</tbody>
</table>

To ensure face validity, each case study was reviewed by an expert clinician in the field or an academic working with that particular client population. Clinicians and academics were invited to read the case study and provide feedback regarding how accurately it represented a client with that condition. Minor factual and grammatical changes were made accordingly (for example the name and type of splint worn by the person in the burn case study was added). Case study number 1, from the first rating session, is attached as Appendix 2.

The case studies were scored twice by participants to establish scale intrarater reliability. For the second collection of data, the case studies were altered slightly to ensure scores were attributable to the reliability of the outcome measure and not the participant’s memory of their scores. Alterations to the case studies included: name, sex and age (by up to 6 years only), and the order of presentation of case studies was reversed. The condition names of seven case studies were changed; however, details of the condition and client presentation remained the same. Table 2 summarises the conditions that were changed.

Procedures

Ethics approval was sought and obtained from the La Trobe University Faculty Human Ethics Committee, and consent was obtained to conduct research at the Donvale Rehabilitation Hospital. Participants attended a 4-h training session, conducted by the second author, that covered the theoretical basis for outcome assessment, and detailed instructions on using and scoring the AusTOMs-OT scales. For data collection, participants
were required to read and score 15 written case studies on the AusTOMs-OT scale of self-care on two separate occasions. The first round of case studies was delivered 2 weeks after the training session to enable participants’ time to practise using the AusTOMs-OT scales clinically. The second set of case studies was delivered 3 weeks after the first round. The optimal time for retesting of data for intrarater reliability is 2 to 3 weeks as this reduces the likelihood that participants will remember their scoring of the first session (Kaplan, 1996). Data were also collected about participants in the form of an anonymous questionnaire asking: their age, years of experience as an occupational therapist and qualifications.

Data analyses
Data were analysed according to the generalisability theory (G) (Crook & Algina, 1986; Streiner & Norman, 1995) and the classical test theory, using an intraclass correlation coefficient (ICC, type 2,1) Shrout & Fleiss, 1979). The ICC determines if scores between raters do not only correlate but that there is agreement in the scores (Portney & Watkins, 2000). The generalisability theory is related to but expands on the classical test theory to include multiple sources of error and gives a composite overall indication of reliability. For the purpose of this study a correlation coefficient of 0.90–1.0 was considered very high, 0.7–0.90 high, 0.50–0.70 moderate and 0.30–0.50 low (Hinkle, Wiersma & Jurs, 1998).

Interrater reliability
Ratings of the case studies on each occasion were compared between participants to calculate interrater reliability coefficients for each domain of the self-care scale using both the generalisability theory and the classical test theory. The generalisability theory gives an overall reliability coefficient (G); however, the classical test theory provides separate ICCs for each rating session. As well, using the generalisability theory, data collected from both rating sessions for all raters were used to expand on traditional interrater reliability by determining reliability of the AusTOMs-OT self-care scale both between raters and over time. The result (G) is a calculation of the reliability of the measure from one observer on one day to another observer on a different day. This mimics the clinical situation where one therapist does an initial assessment and transfers the client to another therapist who completes the follow-up assessment.

Intrarater reliability
Participant ratings of the case studies from the first session were compared with ratings from the second session to establish the intrarater reliability of each domain of the AusTOMs-OT self-care scale using both the generalisability theory and the classical test theory. Again, the classical test theory provides separate ICCs for each rater, whereas the generalisability theory provides an overall measure of intrarater reliability (G). To ease comparison the ICCs were averaged using Fisher’s $r$ to $r'$ method (Howell, 1992).

Retest reliability in the units of measurement
Finally, to provide a method of interpretation of results, retest reliability in the units of measurement was calculated. This determines how much change would need to occur on the AusTOMs-OT self-care scale for that change to be regarded as real change for the client with 95% confidence. As the case studies did not change between rating sessions, the difference between the first and second scores indicates rater error. Thus, real change must exceed that of rater error. This error was used to calculate the 95% confidence intervals. The results were tabulated for groups and for an individual for each rater. Group results are important for research or program evaluation, whereas clinicians need to be confident about change for an individual client.

Table 2: Details of condition names changed from the first to second case studies

<table>
<thead>
<tr>
<th>First round case study no.</th>
<th>Original condition title</th>
<th>Changed condition title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chronic obstructive airways disease</td>
<td>Emphysema</td>
</tr>
<tr>
<td>5</td>
<td>Amputation at ankle</td>
<td>Symes amputation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(amputation at the ankle joint)</td>
</tr>
<tr>
<td>8</td>
<td>Acute myocardial infarction</td>
<td>Heart attack</td>
</tr>
<tr>
<td>9</td>
<td>Cerebrovascular accident</td>
<td>Stroke</td>
</tr>
<tr>
<td>10</td>
<td>Acquired brain injury</td>
<td>Head injury</td>
</tr>
<tr>
<td>11</td>
<td>Dementia</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>15</td>
<td>Chronic kidney disease</td>
<td>Renal failure</td>
</tr>
</tbody>
</table>
Results

Interrater reliability of the AusTOMs-OT self-care scale

Scores for each case study gathered in the first and second rating sessions were compared for the same therapist and were used to calculate interrater reliability coefficients for each domain. Using the generalisability theory, the overall interrater reliability coefficients were 0.74 for the impairment domain, 0.88 for activity limitation domain, 0.81 for participation restriction and 0.94 for the distress/well-being domain. Table 3 presents the interrater reliability intraclass correlation coefficients from the classical test theory. The intrarater reliability coefficients calculated using the generalisability theory were high to very high, whereas those calculated using the classical test theory resulted in moderate to high intrarater reliability. A number of raters had low reliability for the impairment domain, which is represented with the average of 0.55. Rater two appeared to have coefficients well below the average for the activity limitation and participation restriction domains, and was below average for the distress/well-being domain. The reliability of rater 6 was also below average for the activity limitation, participation restriction and distress/well-being domains.

Retest reliability in the units of measurement

Data from both sessions from all raters were analysed to calculate the retest reliability in the units of measurement. This resulted in a confidence interval representing the amount of change in a client’s score that would need to occur to be 95% confident that real change had occurred. The confidence intervals are presented in Table 5, where the lower score represents the amount of change needed to be confident the client’s condition had deteriorated, and the higher score indicating the amount of change in AusTOMs-OT self-care scale needed to be confident the client had improved. The

### Table 3: Self-care scale Interrater reliability coefficients from the generalisability theory (using r statistic) and the classical test theory (using intraclass correlation coefficient (ICC) statistic)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Generalisability theory (G)</th>
<th>Classical test theory (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>Impairment</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>Activity limitation</td>
<td>0.87</td>
<td>0.85</td>
</tr>
<tr>
<td>Participation restriction</td>
<td>0.86</td>
<td>0.87</td>
</tr>
<tr>
<td>Distress/well-being</td>
<td>0.91</td>
<td>0.93</td>
</tr>
</tbody>
</table>

### Table 4: Self-care scale intrarater reliability intraclass correlation coefficients (ICCs) from classical test theory

<table>
<thead>
<tr>
<th>Rater</th>
<th>Impairment</th>
<th>Activity limitation</th>
<th>Participation restriction</th>
<th>Distress/Well-being</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>0.39</td>
<td>0.80</td>
<td>0.77</td>
<td>0.96</td>
</tr>
<tr>
<td>Rater 2</td>
<td>0.25</td>
<td>0.47</td>
<td>0.39</td>
<td>0.65</td>
</tr>
<tr>
<td>Rater 3</td>
<td>0.91</td>
<td>0.67</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>Rater 4</td>
<td>0.75</td>
<td>0.63</td>
<td>0.80</td>
<td>0.66</td>
</tr>
<tr>
<td>Rater 5</td>
<td>0.25</td>
<td>0.87</td>
<td>0.54</td>
<td>0.83</td>
</tr>
<tr>
<td>Rater 6</td>
<td>0.32</td>
<td>0.66</td>
<td>0.65</td>
<td>0.77</td>
</tr>
<tr>
<td>Rater 7</td>
<td>0.55</td>
<td>0.64</td>
<td>0.59</td>
<td>0.90</td>
</tr>
<tr>
<td>Average</td>
<td>0.55</td>
<td>0.70</td>
<td>0.71</td>
<td>0.85</td>
</tr>
</tbody>
</table>
data presented in Table 5 suggest that real change for a group of clients is indicated by a decrease or increase of 1 point for the impairment domain, a decrease of half a point or an increase of 1 point for the activity limitation and participation restriction domains, and a decrease or increase of half a point for the distress/well-being domain. For individual clients, confidence of real change occurs with a decrease of 2.5–3 points or an increase of 2.5 points for the impairment domain, a decrease of 1.5 points or an increase of 2 points for the activity limitation domain, a decrease or increase of 2 points for the participation restriction domain, and a decrease or increase of 1.5 points for the distress/well-being domain. Change within these ranges could be because of rater error rather than real client change.

Discussion

Interrater reliability of the AusTOMs-OT self-care scale
The interrater reliability coefficients for the AusTOMs-OT self-care scale were consistent between the generalisability theory and the classical test theory with high to very high interrater reliability with little variation occurring between the four domains. The results indicate that all domains of the AusTOMs-OT self-care scale demonstrate consistent scores for clients when scored by multiple raters on the same occasion with relatively little error between raters. Data collected with the AusTOMs-OT self-care scale can therefore be pooled to create large quantities of data suitable for use in research that evaluates the efficacy of occupational therapy intervention. This is supported by Eakin (1989). The more occupational therapists who use the same assessments, the easier it becomes to evaluate and compare treatment methods and outcomes. Without such measures, occupational therapists can never prove their true worth’ (p. 53).

Examination of reliability from one observer on one day to another observer on a different day
The participant ratings of the case studies were generalised to create reliability coefficients that represented the reliability of scores both between raters and over time. The reliability coefficients suggest that results from the AusTOMs-OT self-care scale can be generalised between raters and over time with high reliability for the impairment, activity limitation and participation restriction domains, and very high reliability for the distress/well-being domain. These results suggest that the AusTOMs-OT self-care scale has sufficient reliability to evaluate change both over time and between therapists. This is a particular advantage when clients may be rated by different therapists pre- and post-interventions such as when moving from acute to subacute care. The results also support the idea that data collected with the AusTOMs-OT self-care scale can be used for outcomes research to test the effectiveness of occupational therapy intervention.

Intrarater reliability of the AusTOMs-OT self-care scale
The intrarater reliability of an outcome measure provides information on the tool’s ability to accurately measure change over time (Ottenbacher, 1997; Portney & Watkins, 2000; Streiner & Norman, 1995). The intrarater reliability coefficients of the AusTOMs-OT self-care scale showed more variation between the generalisability theory (high to very high reliability) and the classical test theory (moderate to high reliability) than observed for interrater reliability. Despite this variability, the coefficients for the activity limitation, participation restriction and distress/well-being domains have high intrarater reliability, suggesting that they can accurately measure client change over time. The coefficients for the impairment domain, however,
represent only moderate to high levels of intrarater reliability. The lower reliability of the impairment domain may be because of factors within the rater such as experience; however, as participants’ demographic data were anonymous, such comparisons cannot be made.

Demonstrating change in client function provides data that can be used to attribute outcomes to intervention (Foto, 1996; Law & Baum, 1998; Lloyd & King, 2002). It is important for occupational therapists to demonstrate that intervention is cost-effective and provides positive outcomes for clients (Fricke, 1993). The results suggest that the AusTOMs-OT self-care scale can identify real change in client function, which ultimately enable therapists to link client outcomes with intervention.

Retest reliability in the units of measurement

Reliability in the units of measurement refers to the amount of change in a score that would need to occur to be 95% confident that real change occurred for the client. In other words, it is important to understand whether change in a client’s score on the AusTOMs-OT self-care scale is because of real change or unpredictable variability associated with the measurement process. The reliability was calculated for individuals and groups because clinicians are concerned with individual performance and also evaluate services and group programs, and researchers are usually more concerned with group results. It was found that to be 95% confident of real change using the AusTOMs-OT self-care scale, less change in a score is required for groups than for individuals (refer to Table 5).

To further appraise the results, the amount of change needed to indicate real change must be compared with the change in a score that is expected. Perry et al. (2003) reported that the mean change in scalar points in clients over time was: impairment: 2.73 (SD 0.95); activity limitation: 2.42 (SD 0.93); participation restriction: 2.76 (SD 1.08); and distress/well-being: 3.12 (SD 1.17). Therefore, the results suggest that the AusTOMs-OT self-care scale could detect change in groups over all four domains, and individuals for the activity limitation, participation restriction and distress/well-being domains as the expected change is outside the 95% confidence intervals. However, observed changes in an individual on the impairment domain of the self-care scale may be obscured by measurement error. In other words, a client would need to change by more than the average expected change for the clinician to be 95% confident that real change had occurred. This result is consistent with the finding in the previous section that coefficients in the impairment domain only represented moderate to high levels of intrarater reliability.

Overview of the reliability of the AusTOMs-OT self-care scale

In this study, the impairment domain of the AusTOMs-OT self-care scale was consistently the least reliable and the distress/well-being domain the most reliable. The activity limitation and participation restriction domains were always similar with high reliability. Occupational therapists use occupation with the aim of facilitating client occupational performance, which directly impacts on a client’s activity and participation and is known to improve overall well-being (American Occupational Therapy Association, 2001; Christiansen, Backman, Little & Nguyen, 1999; Law & Baum, 2001; Law et al., 2002; Law, Steinwender & LeClair, 1998). Occupational therapy intervention does not always directly target impairment and the ICF (WHO, 2002) recognises that treatment of impairment does not always improve activity, participation, or client distress/well-being. It is likely that occupational therapists are more skilled at recognising and rating the domains of activity, participation and distress/well-being as opposed to impairment.

It is also possible that the impairment domain of the AusTOMs-OT self-care scale may not be as clearly defined as the other domains, hence accounting for variation in these reliability scores. Operational definitions can contribute to the reliability of an outcome measure by providing precise instructions for administration and scoring (Eakin, 1989; Fricke, 1993). If they are unclear, reliability may be compromised as interpretation is left to the reasoning of the therapist. The findings suggest that the operational definitions are satisfactory for the AusTOMs-OT self-care scale. In addition, it is possible that impairment is easier to assess when the full extent of injury can be seen or physically examined by therapists. The written information provided in the case studies may not have been adequate for participants to identify the level of the case study patient’s impairment, and thus have contributed to lower reliability results for the impairment domain.

Limitations and directions for future research

All participants in the study were from the same physical rehabilitation facility treating mainly older persons. Hence, the participants may have been more skilled at rating the older client case studies and less adept at rating the case studies involving younger clients or clients with a mental health diagnosis. Also, the participants with fewer years experience have been less skilled at rating the variety of case studies because of lack of exposure to such clients. In such instances the reliability coefficients obtained may have
be present. However, occupational therapy education necessitates that participants gain an understanding of these clients and so a basic understanding would be present.

The use of written case studies may also have affected the results as therapists would normally administer the AusTOMs self-care scale with clients directly. However, it is difficult to determine whether this would strengthen or weaken the reliability coefficients obtained. In some instances, having written client descriptions replicates the descriptors of the scale too closely, thus leading participants to a rating. In other instances, there may not be enough information resulting in participants guessing the client’s score. The use of written case studies may affect the generalisability of the results to the population if they were not true to life. However, to maximise the face validity of the written cases, experts in the relevant areas reviewed each one. Rating videotaped or real clients is an alternative technique; however, such methods may not provide enough background material. A combination of video and written background material could be utilised; however, there could be more chance that the participants would remember the case clients and consequently their ratings.

Future research is required to conduct detailed reliability studies with the other 11 AusTOMs-OT scales. Furthermore, ongoing studies on the validity, particularly content and construct, are required. While these results and preliminary data published by Perry et al. (2004) suggest that the AusTOMs-OT is reliable and valid, the use of the tool would be enhanced with production of further evidence concerning its psychometric properties.

Summary and conclusion

Overall, the AusTOMs-OT self-care scale was found to have high to very high reliability, suggesting that it has sufficient reliability for use in accreditation, accountability and quality assurance purposes, as well as being suitable for data collection that can contribute to outcomes research. The interrater reliability of the AusTOMs-OT self-care scale suggests that all domains of this scale can reliably score clients observed by multiple raters, and supports the conclusions made by Perry et al. (2004). It can also be concluded that when scoring the activity limitation, participation restriction and distress/well-being domains of the AusTOMs-OT self-care scale, therapists can reliably detect client change. The moderate intrarater reliability of the impairment domain, both in terms of reliability coefficients and expressed in the units of measurement, suggests that this domain has sufficient reliability to detect changes in groups, but should be used with caution for individuals.

Occupational therapists, managers and third party payors can be confident that data collected using the AusTOMs-OT self-care scale has a high level of agreement between multiple raters and that change in scores is because of real client change and not measurement error. Therefore, they have a reliable method of collecting outcome data. Occupational therapists can use outcome data from the AusTOMs-OT self-care scale in research using designs such as randomised controlled trials (RCTs). Such research may aim to attribute client change to the intervention provided and not natural recovery or other maturation effects.

In conclusion, the AusTOMs-OT self-care scale can be used by occupational therapists to gather reliable client data. As such, the AusTOMs-OT self-care scale could be used in research either as the outcome measure used to assess participants in order to investigate or compare intervention; or data collected with the AusTOMS-OT self-care scale could form the subject of research analysis. The outcomes of such research could eventually contribute to the evidence-based practice of occupational therapy.

Acknowledgments

Many thanks to the occupational therapists at Donvale Rehabilitation Hospital for participating in the research. This study was completed in partial fulfillment of the first author’s Honours research project (Fiona Scott, nee Wiseman). This study follows on from the ‘AusTOMs: Australian Therapy Outcome Measures project’ commissioned by the Commonwealth Department of Health and Ageing, Canberra, Australia, and undertaken by the following team members from La Trobe University Faculty of Health Sciences: Professor Alison Perry (Principal Investigator), Professor Meg Morris (Co-investigator), Associate Professor Carolyn Unsworth (Co-investigator), Professor Stephen Duckett (Co-investigator), Ms. Jemma Skeat, Dr Karen Dodd, Dr Nicholas Taylor, Ms Karen Reilly and Ms Dianne Duncombe (Research Associates). Professor Pam Enderby and Dr Alexandra John from Sheffield University, UK, were Associate Researchers on the AusTOMs project.

References


**Appendix 1**

*AusTOMs-OT Self-care scale* (Unsworth & Duncombe, 2004).

Reprinted with permission of La Trobe University.

**Self-care scale**

Self-care consists of: washing and drying body, caring for one’s body (e.g. cutting nails), toileting, grooming (e.g. shaving, brushing hair, applying make-up, cleaning teeth), dressing and undressing, eating and drinking, and looking after one’s health (e.g. taking medication).

**Impairment of either structure or function (as appropriate to age):**

Impairments are problems in body structure (anatomical) or function (physiological or psychological) as a significant deviation or loss. Impairments may be mental (cognitive/perceptual), sensory, voice/speech, cardiovascular/respiratory, digestive/metabolic/endocrine systems, genitourinary/reproductive, neurological movement, or musculoskeletal.

A variety of impairments may impact on the ability to engage in self-care. Considering all the impairments an individual may have, assess the level of severity of these. Base your assessment on typical presentation of the individual’s impairment/s in an appropriate environment.

0. The most severe presentation of impairment/s, e.g. very dense hemiplegia or severe fixed contractures, or constant and intrusive hallucinations or unbearable pain or most severe presentation of cognitive impairment.

1. Severe presentation of this impairment/s, e.g. dense hemiplegia, or severely restricted range of movement or very frequent and intrusive hallucinations or severe pain or severe cognitive impairment.

2. Moderate/severe presentation/s, e.g. moderate to severe hemiplegia, or moderate to severely restricted range of movement or frequent and intrusive hallucinations or moderate to severe pain or moderate to severe cognitive impairment.

3. Moderate presentation/s, e.g. moderate hemiplegia, or moderately restricted range of movement or somewhat frequent but rarely intrusive hallucinations or moderate pain or moderate cognitive impairment.

4. Mild presentation/s, e.g. mild hemiplegia, or mildly restricted range of movement (e.g. morning stiffness) or infrequent and non-intrusive hallucinations or mild pain or mild cognitive impairment.

5. No impairment of structure or function. All structures and functions intact. No pain.

**Activity limitations (as appropriate to age):**

Activity limitation results from difficulty in the performance of an activity. Activity is the execution of a task by the individual. Assess the individual’s ability to perform the multiple activities involved in self-care. Assess what the client actually does.


1. Severe difficulty in performing self-care activities. Requires maximum assistance to perform self-care tasks. May demonstrate an awareness of the processing required for the activity. Individual may offer minimal movement to assist the carer, or maintain a posture.

2. Moderate/severe difficulty in performing self-care activities. Able to perform self-care tasks with hands on assistance from a carer, or constant verbal prompting. Client can perform some parts of the activity, e.g. thread arms into jumper before carer puts over client’s head.

3. Moderate difficulty in performing self-care activities. Able to perform self-care tasks with verbal prompting or supervision or set-up.

4. Mild difficulty in performing self-care activities. Able to do but lacking in quality or extra time required.

5. No difficulty in performing self-care activities. Able to perform all aspects of self-care activities independently with or without use of aids or adaptive equipment, e.g. raised toilet seat. Completes activities in reasonable time.

**Participation restrictions (as appropriate to age):**

Participation restrictions are difficulties the individual may have in the manner or extent of involvement in his or her life situation. Clinicians should ask themselves: ‘given his or her problem, is this individual experiencing disadvantage?’

0. Unable to fulfil social, work, educational or family roles. No social integration. No involvement in decision-making. No control over environment. Unable to reach potential in any situation.

1. Severe difficulties in fulfilling social, work, educational or family roles. Very limited social integration. Very limited involvement in decision-making. Very little control over environment. Can only rarely reach potential with maximum assistance.
2. Moderately severe difficulties in fulfilling social, work, educational or family roles. Limited social integration. Limited involvement in decision-making. Control over environment in one setting only. Usually reaches potential with maximum assistance.

3. Moderate difficulties in fulfilling social, work, educational or family roles. Relies on moderate assistance for social integration. Limited involvement in decision-making. Control over environment in more than one setting. Always reaches potential with maximum assistance and sometimes reaches potential without assistance.

4. Mild difficulties in fulfilling social, work, educational or family roles. Needs little assistance for social integration and decision-making. Control over environment in more than one setting. Reaches potential with little assistance.

5. No difficulties in fulfilling social, work, educational or family roles. No assistance required for social integration or decision-making. Control over environment in all settings. Reaches potential with no assistance.

**Distress/well-being (as appropriate to age):**
The level of concern experienced by the individual. Concern may be evidenced by anger, frustration, apathy or depression.

0. High and consistent levels of distress or concern.
1. Severe concern, becomes distressed or concerned easily. Requires constant reassurance. Loses emotional control easily.
2. Moderately severe concern. Frequent emotional encouragement and reassurance required.
3. Moderate concern. May be able to manage emotions at times, although may require some encouragement.
4. Mild concern. Able to manage emotions in most situations. Occasional emotional support or encouragement needed.
5. Able to cope with most situations. Accepts and understands own limitations.

**Appendix 2**

*Case study number 1 from the first rating session.*

**Case study 1 — Bill**

**Chronic obstructive airway disease**

**Male, aged 86 years**

- Bill was widowed 11 years ago and lives alone in large house in the outer suburbs. He is very proud of his large landscaped gardens.
- Bill is ready for discharge from hospital after a 1-week admission due to a chest infection. On admission to hospital he would become short of breath with any mild physical activity. He would then require oxygen. He had not been participat-