This is the published version of a paper presented at the Graduate Certificate in Tertiary Education Mini-Conference held at The Flinders University of South Australia, Bedford Park, South Australia, November 15, 2000.


Copyright © The Flinders University of South Australia, 2001. Copies available from: School of Education, Flinders University.
A KNOWLEDGE QUESTIONNAIRE ON AMPHETAMINE RELATED PROBLEMS: ASSESSING THE CONCEPTUAL VALIDITY OF THE SCALE, AND THE IMPLICATIONS OF THE METHOD FOR INFORMING TEACHING AND STUDENT LEARNING

Jodie Shoobridge

NCETA
Abstract

Whilst knowledge and attitude scales are frequently used evaluation ‘tools’ in the alcohol and other drug field, rarely do reports exist regarding the conceptual coherence or statistical validity of instruments in common use. Utilising Rasch modelling, a knowledge questionnaire on amphetamine related issues was found to be statistically valid, a finding that supports its conceptual validity. The analysis also provided opportunities for informing the development of future training strategies. The implications for influencing student learning through analysis of individual performance are discussed.

Introduction

Whilst considerable effort has been undertaken in developing and evaluating alcohol and other drug training strategies within the medical profession, rarely are there published accounts of good quality evaluations of training programs for generalist workers (Allsop, Cormack, Addy, Ashenden, Ask, & Beel, 1998; Pols, Cape, & Ashenden, 1993; Bagnall & Fossey 1997; Prior 1994; Roche, 1998; Rutz, von Knorring, Walinder, & Wistedt, 1990). Poorly defined theoretical frameworks for evaluation (Albery, Durand, Heuston, Groves, Gossip and Strang, 1997; Gorman, 1993; Walsh, 1995), poor relationships between evaluation tools and training outcomes (Allsop et al., 1998 Krien & Weldon, 1994; Warr, Allen & Birdi 1999), and ‘spurious’ causal links between the training aims and workplace performance (Carroll, 1995; Cartwright, 1980; Gorman, 1993; Lightfoot & Orford 1986) are among the criticisms of published work in this area.

Knowledge questionnaires play a significant role in the training evaluation controversy (Bagnall & Fossey 1997). They are popular due to their familiarity, ease of administration and marking, and broad ‘knowledge sampling’ potential that is, they can cover various topics in a relatively short time frame. Some argue Jansen, Tan., van der Vleuten, van Luijk, Rethans, & Grol, 1995; Roche, 1996), that within the context of a broader evaluation strategy assessment of performance on specific knowledge areas can provide valuable information. More specifically, it is argued that knowledge gain may reflect a participant’s perception of role adequacy (the ability required to implement new knowledge into future practice (Bush 1987)).

Nevertheless, knowledge questionnaires have attracted much criticism courtesy of their design, for example, some have questioned their ability to measure changes in knowledge, suggesting that these tests merely encourage rote learning (Bramley, 1996, Murray-Harvey, Silins & Orrell, 1996) or simply measure recall (Ewan & Whaitie, 1983). Concerns have similarly been raised regarding the reliability and validity (Bagnall & Fossey 1997) of many of the instruments or scales cited in published literature.

A number of factors contribute to the failure of instruments or tools to measure those things they purport to measure. A few of these are specifically related to instrument design, such as poor validity of the scale (or specifically, the items within the scale); others relate to inadequate sample sizes; sampling variability; and poor scale coherence. Good validity and reliability, combined with relevance, balance, specificity, discriminability and difficulty are considered important properties of a good test (Barnard, 1999; Haccoun & Hamitiaux, 1994). Similarly, a test is only valid for the purposes for which it is defined (Barnard, 1999); in this instance, an assessment of understanding of amphetamine related issues.

Item Response Theory (IRT) provides the theoretical framework for investigating the psychometric properties of this knowledge scale (Keeves & Alagumalai 1999; Waugh, 1999). The fundamental principles underlying this approach aim to ‘improve the measurement properties of total scores by developing an interval scale that is independent of both the items and the persons who have responded to a particular test or attitude scale’ (Keeves & Alagumalai, 1999, p.11) and has the potential for ‘improving the measurement of the abilities or performance of persons using their responses to a set of items’ (Keeves & Alagumalai 1999,p.12). Among the more recent developments in measurement techniques are those deriving from education. Rasch modelling techniques can convert Likert scales and dichotomous scores into linear measures, hence specifying linearly the statistical interaction between people, tests, items and tasks from ordinal data (RASCH 2001). This method can provide information that meets summative assessment requirements (depicting overall scores, determining scale reliability, and evidence of knowledge gain of students partaking in the assessment). In addition, such methods may determine the
appropriateness, or relative difficulty of specific items within a questionnaire, and identify individual performance or ability relative to overall item difficulty (Adams and Khoo 1993).

This methodology also enables the trainer to respond effectively to an evaluation task – to delete inappropriate items within a knowledge scale, identify items too easy or too difficult for similar cohorts of students, to concentrate on specific content areas, or to identify particular areas of difficulty for individual students. The benefits for students resulting from IRT methodology include a demonstration of the student's ability relative to others in the group, or identification of content areas that were not well understood.

A review of the literature has identified that rarely have alcohol and other drugs training evaluations incorporated sophisticated training evaluation methods (Bagnall & Fossey 1997), although a few publications exist that employ Rasch methodology to examine clinical instruments (e.g., see Cornel, Knibbe, Van Zutphen & Drop 1994; Kan, Bretele, Van der Ven & Zitman, 1998). This may be due in part to the relatively recent shift of alcohol and other drugs education and training to the tertiary sector in Australia, changes in perceptions about appropriate training and evaluation design in recent years (Allsop et al. 1998) and increasing interest within the field to formalise training evaluation strategies (Albery et al. 1997), as is intended here. This pilot study therefore aims to contribute to the field through the investigation of the psychometric properties, and the potential to influence teaching and learning, of a knowledge scale on amphetamine related problems (Pead, Lintzeris & Churchill, 1996).

**Method**

**Study design**

The present study is a small part of a larger training evaluation strategy (pre-test post-test design) on amphetamine related problems (Shoobridge 2000). The purpose of this study was to investigate the psychometric properties of one of the tools utilised in the evaluation, and identify how this tool may assist future training situations, and influence student learning. The research was conducted in several stages.

Firstly, the knowledge items were assessed via Rasch modelling as to their conceptual coherence, that is, the procedure aimed to identify whether the items within the scale represented, (or operated as) a singular, unidimensional concept or construct, about amphetamine related issues. Unidimensionality is a necessary indicator that the scale had good content validity (Barnard 1999)). In establishing the reliability of the scale, information was also obtained on item (content area) difficulty, which can in turn, inform future training activities. Thus, the focus in this part of the study is on the characteristics of the items that constituted the scale.

The efficacy of the intervention, indicated by changes in knowledge that occurred following training was investigated through a comparison of post-test and follow-up scores. The focus in this aspect of the study was on the scaled interval scores that were achieved by each of the persons who participated.

The next task involved identifying how much movement occurred across the items for each individual at each time point, and to identify similarities in response patterns over time. This procedure enabled the trainer to:

1. provide a benchmark of anticipated performance for different groups of students (The underpinning philosophy of IRT is to offer item-free measures of individual performance or ability, and person-free measures of item characteristics (Barnard 1999). As such, similar results should be obtained on each occasion of use.)
2. identify items or related concepts in relation to their difficulty, or relative ease, in order to inform future planning of training events
3. elicit information regarding the performance of individuals
4. determine whether or not individual responses patterns were consistent with the difficulty of the items
5. provide feedback on individual performance across time, and relative to other participants
Setting and participants

Data were obtained from participants who attended a one-day training program on amphetamine related issues. The aim of the workshop was to encourage the development of appropriate responses to clients who were experiencing amphetamine-related problems. The participants represented health and community workers either in a generalist setting, or a drug and alcohol or mental health specialist setting, who had a particular interest in amphetamines. All program participants were invited to complete a series of questionnaires, among them a questionnaire designed specifically to test knowledge about amphetamine related issues (Pead et al, 1996). The evaluation required participants to complete the questionnaires (including the knowledge scale) on three occasions - immediately prior to the workshop (pre-test), immediately following the workshop (post-test), and two months following the workshop (follow-up). Thirty-six participants provided complete information for each of the three evaluation phases. Twelve of these participants were also involved in either face-to-face or telephone interviews, thus providing an additional rich source of qualitative data.

Knowledge Scale

The knowledge scale (Pead et al 1996) consisted of 28 items, eliciting a ‘true’, ‘false’ or ‘don’t know’ response (Appendix 1). An additional question was added that was considered pertinent to test new information added into the training program.

Data analysis

The Rasch scaling analysis was performed through QUEST interactive software (Adams & Khoo, 1993).

Item scalability

The study investigated the Rasch homogeneity of the knowledge scale. A number of assumptions specified in the Rasch model were tested, in order to ascertain whether the Rasch model held true (Kan, Breterler, Van Der Ven & Zitman, 1998). The basic assumptions underlying IRT are described below (Weiss & Yoes, 1991 p. 73):

Firstly, it is assumed that when a person knows the correct answer, they will answer it correctly. It follows that if they answered incorrectly, or checked a ‘don’t know’ response, they did not know the correct answer.

Secondly, the concept of unidimensionality holds that a group of items in a scale together represent, or function as, a singular construct, an underlying or latent trait or variable that may be knowledge, attitude, or ability, (Yen & Edwardson 1999, Keeves & Alagumalai 1999). Hence a person’s performance (the probability of them providing a correct response) is attributable to that particular latent trait (Yen & Edwardson 1999). A scale must satisfy the unidimensionality requirement before analysis can proceed.

Local stochastic independence requires that items are independent of each other, and that each person has a certain probability of giving a predefined response to that item (Ken et al 1998). If the underlying construct or trait is constant, then no two items would provide similar information, that is, a student could not gain clues to the correct response for an item based on an answer to a previous item (Weiss & Yoes, 1991).

An item response function (IRF) describes mathematically the relationship between the item difficulty and a respondent’s ability, and is represented graphically as the item characteristic curve (ICC) (Weiss & Yoes, 1991). An IRF relates the probability of a person correctly answering an item to the person ability parameter ($\theta$). On a knowledge scale, for example, there are likely to be a number of levels of ability to answer the questions correctly. Each of these levels are represented by an estimate of the probability that someone with a certain level of ability will correctly answer a given item. Equation 1 shows that where $\theta$ is the ability of person $i$, $b_g$ is the difficulty of item $g$, and $P_g(\theta)$ is the probability that person (ability $\theta$) will correctly answer item $g$. $e$ is the natural logarithm base (Yen & Edwardson 1999).

$$P_g(\theta) = \frac{e^{(\theta_i - b_g)}}{1 + e^{(\theta_i - b_g)}} \quad (1)$$
According to the one parameter Rasch model (used when items are described only in terms of their item difficulty), the difficulty, or threshold, of the item is the probability of the person having a 50% chance of answering the item correctly ($P(g|\theta) = 0.5$). The harder the item, the more ability the person must have to answer the question correctly ((Weiss & Yoes, 1991; Yen & Edwardson 1999).

**Scale discriminability**

The utility of the scale depends on the discriminative power (or fit) of both items and subjects. Scale discriminability refers to the ability of the scale to discriminate amongst people with different levels of ability (in this case, varying levels of knowledge).

**Construct validity**

Construct validity requires that there is a theoretical rationale (Kan et al 1998) underlying the knowledge scale on amphetamine related issues. The design of the training package (and assessment tasks) incorporated input and critique from clinicians familiar with the content (Pead et al 1996), and the final version is purposed to have good face validity.

**Results**

The health workers had a median age of 40 years (range 21-55), and comprised 9 males and 27 females. Six of the workers (17%) were employed at a drug specialist agency. Forty eight per cent of the sample stated that they had at least a ‘reasonable amount’ of experience in working with amphetamine users.

In order to examine whether changes in knowledge were detectable for the group prior to Rasch scaling the questionnaire (Pead et al 1996, Appendix 1), a repeated measures analysis of variance was performed. There were significant differences in levels of knowledge at each time point ($F(1) = 37.87, p<.001$, partial $\eta^2 = .52$), with planned comparisons showing these changes to be significant at both post test and 2 month follow up. Table 1 shows the median and mean knowledge scores for the sample at each time period.

<table>
<thead>
<tr>
<th>Table 1. Median Knowledge scores</th>
<th>Median/29</th>
<th>Mean (s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>13</td>
<td>13.94 (5.5)</td>
</tr>
<tr>
<td>Post-test</td>
<td>20</td>
<td>19.25 (3.8)</td>
</tr>
<tr>
<td>Follow-up</td>
<td>19.5</td>
<td>19.5 (4.9)</td>
</tr>
</tbody>
</table>

**Item treatment**

Prior to the commencement of data analysis, the item responses ‘true’, ‘false’ and ‘don’t know’ were recoded into dichotomous responses. A ‘true’ response was recoded to represent a correct response, whereas ‘false’ and ‘don’t know’ responses were recoded to represent an incorrect response.

The first task involved the calibration of the items within the scale to test the fit of the estimates obtained from the item parameters to the Rasch model. As this was a repeated measures design with relatively small sample size, the calibration of the scale was enhanced by combining the post-test and follow-up scores (D. Curtis, Flinders University of SA, personal communication, January 16, 2001). The calibrated scale was found to satisfy the assumptions of unidimensionality (Mean Infit MS=1.00, s.d=0.22; Outfit MS range=0.8-1.18). Overall, the scale met the assumptions of local independence and mathematical function, by demonstrating that the items were similar enough to represent a coherent body of knowledge of amphetamine related issues (Figure 1). Figure 1 shows the fit of the items (between the dotted lines) according to their estimated scores, within acceptable range (.77 – 1.33).
<table>
<thead>
<tr>
<th>INFIT</th>
<th>MNSQ</th>
<th>0.63</th>
<th>0.71</th>
<th>0.83</th>
<th>1.00</th>
<th>1.20</th>
<th>1.40</th>
<th>1.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 item 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 item 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 item 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 item 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 item 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 item 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 item 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 item 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 item 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 item 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 item 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 item 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 item 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 item 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 item 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 item 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 item 17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 item 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 item 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 item 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 item 21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 item 22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 item 23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 item 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 item 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 item 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 item 27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 item 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29 item 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Graphical representation of the amphetamine knowledge items post calibration

Whilst an item analysis revealed poor point biserial correlations for five items (items 1, 13, 18, 26 and 29), the Rasch analysis overall revealed a good fit for the items. The removal of these items did not substantially improve the existing internal reliability, and the results obtained from the Rasch analysis suggest that the items in question carry little information value and might be revised. However, their acceptable fit statistics suggest that they introduce no bias in the overall scale, and thus have been retained. Because of the good item fit parameters, the scale appears to demonstrate good construct validity. According to classical test theory, the questions formed a reliable scale (Cronbach’s alpha= 0.90).

The next step involved verifying that the instrument measured changes in subject performance following the training (hence demonstrating that the intervention was effective). At this point, the item estimates were ‘anchored’, that is, the items were set in their calibrated position in order remove variation due to subject performance at the following time points. (This procedure enables differences in subject scores at subsequent time points to be detected). Both the raw post-test scores and follow-up scores were separately analysed against the calibrated scale. Post-test case (person) estimates revealed acceptable fit statistics (Mean Infit MS=.96, s.d.=.19) and revealed an overall consistent pattern of participant responses to the items. Figure 2 is a graphical demonstration of the estimates of the individual scores (indicated by X) against the calibrated scale at post-test. Each item (Appendix 1) can be identified on this scale by their item number. (The items have been calibrated in decreasing order of difficulty. An item placed at the 0 point on the scale represents average difficulty - the level at which you would expect people with average ability to answer this item (and those below it), correctly, whereas item 11 is the most difficult item).

Overall, participants demonstrated increases in performance relative to the calibrated items at post-test, performing with at least average ability, as would be expected should the intervention be considered successful. A slight performance drop was detected for 3 subjects, however, this decrease was not significant (within S.E. of >5 or less).

At follow-up, most participants performed as expected (Mean Infit MS=1.02, s.d.=.23), demonstrating that their scores were increased or maintained from their post-test scores (Figure 3).
However several aberrant scores were detected through an examination of case estimates, and are
visually represented in Figure 3. Four participants scored either much higher than expected (with a
score well above the most difficult item on the scale (item 11, with estimates at 3.84), or much lower
case estimates relative to post-test results (scoring well below the expected range, with a case
estimate of -0.99). Examination of the case estimates suggest that these participants did not perform
within expected range (indicated by large standard errors).

---

**Figure 2.** Quest Output: Representation of participant performance (indicated by X) against
calibrated items (indicated by item number) at post-test.
Figure 3. Quest Output: Representation of participant performance (indicated by X) against calibrated items (indicated by item number) at follow-up.

From Figure 4, it can be seen that the estimates for participants 25, 27 and 30 at follow-up were almost 4 logits away from the mean (with a SE considerably larger than the expected range of 0.4 - 0.5). In contrast, subject 14, performed considerably less well at follow-up. The results obtained for subject 2, who performed as expected (with a consistent SE), has been randomly selected and included in the graph for comparative purposes.
Further analysis (not shown here) utilising Rasch modelling provides additional information regarding individuals' performance relative to the items. Comparison of raw and estimated scores, and graphical representation (through the production of kidmaps) may be elicited through QUEST to assist students to identify their performance relative to each item.

**Discussion**

This pilot study has demonstrated the use of Rasch modelling for the purposes of calibrating a knowledge scale in common use for training in the drug and alcohol field. This study has provided information regarding the validity and reliability of the scale, the performance of the items that comprised the scale, and information that may enable teaching and learning strategies, through identification of difficult and relatively easy items (to assist in planning and conducting training). Student performance may be then assessed relative to item difficulty.

The knowledge scale was found to represent a coherent body of knowledge of amphetamine related issues. In addition, the scale demonstrated a wide range of difficulty, hence useful in discriminating over a wide range of abilities. In calibrating the scale, a number of items initially appeared to be a poor fit for the model. However, the over-inclusive nature of each of these items suggested that the retention or removal of these items was arbitrary. Trainers interested in further tailoring their training could delete these items without affecting the validity and reliability of the scale.

Examination of the placement of the items (against subject performance) during calibration may provide valuable information with which to guide trainers in identifying the more conceptually difficult items, and those items that were relatively easy (that participants should be expected to answer correctly following a training program). Utilising the information provided by QUEST through Rasch modelling may assist trainers in future program planning in order to ensure that adequate coverage of conceptually difficult topics occurs, or that the program is tailored to particular groups that may not require such depth.

For example, significant time is granted within the training package for the content regarding question 11, the most difficult item in the scale following calibration. Group performance on item 11, 'Amphetamine users seeing general practitioners are usually seeking depressant drugs' suggested that this item following training and additional experience, this item, amongst others, remained difficult to answer. The training package (and training conducted) placed considerable emphasis on this concept during training (to encourage workers to examine the view that these clients are as likely as anyone to present for a whole range of problems). At follow-up only 5 participants were able to answer this item correctly. Examination of the concepts related to this item in a training situation, and participants' experiences in this area, could be further examined in future training situations.

Rasch demonstrated knowledge increase at post-test, which, for most participants, was at least maintained at follow-up. However, Rasch modelling extends traditional feedback methods by providing specific information to individuals about their actual performance beyond that of providing an overall score. Rasch modelling provides information regarding an individual's knowledge relative...
to other participants, and about an individual’s performance relative to specific items, a situation unique to this methodology, and IRT.

The ability of Rasch modelling to identify aberrant scores may be particularly useful, in flagging issues about an individual’s performance that may be quite unrelated to the training situation. For example, an individual’s poor performance may be indicative of fatigue with the repeated administration of the instrument, that they were preoccupied with other matters, or they were anxious to leave and did not attend closely to the demands of the task. Performance at (especially) long term follow-up may reveal other information may indicate areas of further exploration or research. For example, many work related factors may influenced poor follow-up performance, such as workplace distractions, reluctance or unwanted pressure to repeat the test, attempts at completion in order to attain the incentives that may be provided (hence they may have little concern for answering correctly). It would also be anticipated that over time, workplace factors (returning to the same environment, lack of workplace changes consistent with training), may influence a return to familiar routines, hence possibly explaining to some degree, knowledge fade over time. This information may be invaluable in terms of informing the development of post-training workplace support strategies.

Alternatively, several participants performed considerably better than expected (relative to post-test case estimate scores and standard error) at follow-up. Again, several explanations are possible, including the possibility that additional experience in working with amphetamine users since the intervention consolidated their knowledge. Other explanations may include that the person scored very well, purely by chance, or that some assistance was available at the time the test was taken. Again, such changes may be indicative of the need for further exploration of factors that may enhance or support knowledge gain in the workplace.

In the past, rarely have trainers in the alcohol and other drug field engaged in ‘high stakes’ testing in a capacity similar to that of the field of education. Similarly, recognition that evaluation methods for training programs are frequently inadequate, that there are increasing demands to justify training budgets and provide measurable outcomes, industrial changes, and increasing workforce demands for evidence of competence suggest that further exploration of the use of different evaluation methods are warranted. Rasch modelling is but one method that may provide significant opportunities for the drug and alcohol field. For example, calibrated scales that represent a coherent body of knowledge, that are context and person free and have proven validity and reliability, provide both trainers and their students with both the evidence, and confidence, that a genuine change has occurred. Similarly, Rasch modelling enables a feedback process to inform and guide student performance relative to the items and other students, or identify areas where students experienced difficulty with certain concepts.

This study was an exploratory study of a knowledge scale using methods grounded in Item Response Theory (IRT) using existing data and a published questionnaire. Utilising concepts grounded in IRT, the project had the dual purpose of strengthening the results of a more comprehensive evaluation of a training program for frontline workers on amphetamine related problems through testing the knowledge scale used in the training program. Rasch modelling enabled the calibration of the scale (to examine whether the items represented a singular concept or trait), to identify concepts that were poorly understood following training, with additional benefits of providing information about individual performance (or ability), relative to the difficulty of the items. Prior to this study, there was no information available regarding the performance and validity of the scale utilised in the study, nor information about the extent that the items reflected a coherent body of knowledge about amphetamine related problems, beyond face validity.

Acknowledgements

I would like to thank David Curtis, Roz Murray-Harvey, Professor John Keeves, and Sharyn Watts for their assistance throughout the research process, and for their critical comment in the preparation of this paper.
References


Barnard, J.J. (1999) Item Analysis in Test Construction (Publication details unknown at time of printing)


Keeves, J. (2000) Personal communication, Flinders University of South Australia

Keeves, J. and Alagumalai, S. (1999) New approaches to measurement (Publication details unknown at time of printing)


Roche (1996) Increasing primary care providers willingness to intervene in alcohol and drug related problems: A review. Substance Abuse 17(4): 201-17


Appendix 1

Amphetamine knowledge questionnaire items

1) Most people will use psychostimulants at some time in their lives
2) Psychostimulants are usually manufactured in backyard laboratories
3) Amphetamines are usually prescribed for overeating, sleep disorders and hyperactivity
4) It is important to determine the chemical name of an amphetamine
5) The effects of ecstasy are similar to those of amphetamine
6) Amphetamine users are primarily young males with a polydrug use history
7) Amphetamine harms may arise from drug administration, intoxication and withdrawal
8) Duration of amphetamine effects are mainly related to the route of administration
9) The amphetamine crash is associated characteristically with bizarre thoughts
10) Amphetamines are not usually used heavily for periods longer than 10-14 days
11) Amphetamine users seeing general practitioners are usually seeking depressant drugs
12) In taking a history an equal emphasis should be given to all drugs used
13) Most amphetamine users know that there are negative effects
14) Harms from amphetamines mainly relate to injecting behaviour
15) Amphetamine related mental state problems may be indistinguishable from schizophrenia
16) Aggressive behaviour may be directly attributable to amphetamine intoxication
17) Confronting the intoxicated person with the effect of their behaviour on others is important
18) Most regular amphetamine users will have experienced mental state problems
19) Vulnerability to amphetamine psychosis is mainly related to family psychiatric history
20) Most toxic amphetamine effects usually respond to cooling, hydration and reassurance
21) The amphetamine withdrawal syndrome may last several months
22) Psychotic features may re-emerge during the withdrawal syndrome
23) Oversleeping, psychomotor retardation and suicidal ideas are the main withdrawal features
24) Dose, duration of use and withdrawal history are the best predictors of withdrawal severity
25) Amphetamine withdrawal should generally be conducted in an inpatient setting
26) Psychotropic medication over several months is usually required in withdrawal treatment
27) Psychosocial treatments for prolonged withdrawal focus on managing cravings
28) Lapses back to amphetamine use usually means that heavy use of amphetamines will ensue
29) Methadone is an appropriate treatment option for heavy users of amphetamines