

Thus it has been used in Holland and in Sweden in connection with attempts to introduce systematic data collection in treatment systems where social work has had a major role (Broekman *et al.* 2004; Tengvald *et al.* 2004). In Hungary, on the other hand, one important aim has been to include social aspects in the problem assessment in a psychiatry-dominated system (Gerevich *et al.* 2004). In Denmark, the introduction was connected with building documentation systems to find 'evidence-based' treatment methods (Vind & Hecksher, 2004). In Norway, development has been cautious and driven by researchers (Lauritzen & Ravndal, 2004). The list of curricula could be made even longer.

The ASI can profit from the status that a US product automatically carries in many environments; but the instrument could not have served as a joker without claiming a scientifically ascertained status. The temptation to exaggerate positive research findings and disregard problematic aspects has been great, particularly where the political and administrative reform demands have been powerful. In addition, as is often the case when ideas diffuse, the followers have sometimes become more papal than the Pope.

Klaus Mäkelä's paper (Mäkelä 2004) makes this status of the instrument a problem, and argues convincingly that 'it is misleading to speak of the reliability and validity of the ASI as an entity... Instead we need a discussion of what measures based on what parts of the ASI can be used in what populations and for what purposes'.

Is the urge to find a quick fix for the problems of the treatment system so great that the critiques of Mäkelä and others will be taken as an argument for abandoning the ASI altogether? Or do our politicians and administrators have the patience to allow analyses and learning from the ASI experience? The latter attitude would be the fruitful one. The Joker has magic qualities, but the ASI can be more useful when transformed into a less exciting, but more precise hand of cards.

- Gerevich, J., Bácskai, E. & Rózsa, S. (2004) Hungarian adaptation of the European Addiction Severity Index. *Journal of Substance Use*, in press.
- Lauritzen, G. & Ravndal, E. (2004) Introduction of EuropASI in Norway: clinical and research experiences from a cost-benefit-study. *Journal of Substance Use*, in press.
- Mäkelä, K. (2004) Studies of the reliability and validity of the Addiction Severity Index. *Addiction*, 99, 398–410.
- Stenius, K. & Room, R. (2004) Measuring 'addiction' in Europe: the diffusion of the Addiction Severity Index (ASI), and its purposes and functions. *Journal of Substance Use*, in press.
- Tengvald, K., Andrén, A., Bergman, H., Engström, C., Nysrötöm, S., Sallmén, B. & Öberg, D. (2004) Implementing the Addiction Severity Index (ASI) in Swedish human services sectors—experiences, problems and prospects. *Journal of Substance Use*, in press.
- Vind, L. & Hecksher, D. (2004) The diffusion of ASI in Denmark. *Journal of Substance Use*, in press.

THE BATH WATER OR THE BABY?

It is, perhaps, always a good idea to draw attention to the intellectual slumber that we can be lulled into by having a 'standard' that everyone uses and no one questions. Klaus Mäkelä (2004) takes a look at the 20-year-old workhorse, the Addiction Severity Index (ASI), and asks if we accept too readily that it is reliable and valid. The ASI is used nationwide by scientists, clinicians, administrators and policymakers to get a handle on the kinds of problems substance abuse clients bring to treatment, and sometimes to gauge changes that result from treatment. Indeed, the ASI may be one of the most successfully disseminated (cf. Budman *et al.* 2003) assessment tools ever developed in the addictions field. Given such widespread use and acceptance, it is critical that we do not use it blindly.

That said, it is important that such a re-examination is informed by a sophisticated understanding of the psychometrics involved, so that current and potential consumers of ASI data are not misled. Mäkelä addresses important sources of measurement error. However, much of Mäkelä's criticism of the psychometrics of the ASI is misleading.

Mäkelä cites uncritically Nunnally & Bernstein's (1994) unfortunate recommendation that 'a reliability of 0.90 is the bare minimum' (p. 265). This statement is controversial at best. Space prevents a full discussion, but internal consistencies this high, in an ASI type of assessment, most probably represent unnecessary redundancy (Streiner 2003a,b). In some cases (see Loevinger's classic 1954 article), increasing coefficient alphas can actually lower validity. In an era where long assessments are not likely to be used, a naive emphasis on very high internal consistency may guide scale developers and users in the wrong direction. Indeed, it is critical to keep in mind that

The definitive version is available at www.blackwell-synergy.com
<http://blackwellpublishing.com/journal.asp?ref=0965-2140&site=1>

KERSTIN STENIUS
Nordisk Alkohol-och Narkotikatidskrift
Stakes
PI 220
Helsinki
Fin-00531
Finland
E-mail: kerstins@stakes.fi

REFERENCES

- Broekman, T. G., Schippers, G. M., Koeter, M. W. J. & van den Brink, W. (2004) Standardized assessment in substance abuse treatment in the Netherlands: the case of the ASI and new developments. *Journal of Substance Use*, in press.

reliability is 'a characteristic of test scores, not of the test itself' (Streiner 2003a, p. 101, italics original). This is true as well for test-retest reliability which, as Ware and colleagues (Stewart *et al.* 1992) point out, is a function of the time interval chosen, the nature of the measure and the characteristics of respondents. Reliability parameters must be realistic and suited to the measure's purpose and the target population.

The discussion of validity coefficients is also superficial. For example, correlations between the ASI alcohol and drug composite scores with such measures as the Michigan Alcohol Screening Test (MAST) and Drug Abuse Screening Test (DAST) may not be particularly high due to problems inherent in the comparison measures, rather than the ASI, or to the different purposes of the ASI and such comparison measures. Cross-cultural comparisons may be more effected by poor adaptations of the measures involved (e.g. Geisinger 1994; Rogler 1999) than qualities of the original measures. Discussions of sensitivity and specificity across different populations should include a discussion of Bayes' Theorem (Meehl & Rosen 1955), which describes how the prevalence of the target condition in a specific population can dramatically affect the predictive value of a measure's cutoff score. Without careful consideration of such issues, conclusions that the problem lies with the ASI, particularly as it is meant to be used, are unwarranted and dubious.

Mäkelä's point on the complexity of the ASI administration and scoring and the need for extensive and expensive training and monitoring of inter-rater reliability is well taken. Turnover at substance abuse treatment centers averages 50% per year (Carise *et al.* 2003). This alone threatens any notion of adequate training. Many authors, including McLellan (TRI 1990), have bemoaned poor interrater reliability problems of the ASI. However, our work on developing a multimedia, self-administered version of the ASI, complete with reliable and valid, algorithm-generated severity scores (Butler *et al.* 1998, 2001), goes some distance toward overcoming this critical source of measurement error.

The ASI has been extraordinarily successful in emphasizing the multi-dimensional nature of problems presented by substance abuse clients and providing a standard assessment for the field. The ASI is not a perfect assessment, but I would argue that it achieves a reasonably good balance of assessment length, reliability and validity. It should not be extended to purposes and populations for which it is not intended without first establishing its appropriateness. Nor should the ASI's reliability and validity be oversold. New revisions and entirely new assessments should be sought continually (although these are likely to be validated against the ASI) and improved administration technologies should be

explored. The concern is that, in the name of rigor, unrealistic or unwarranted reliability and validity values become the benchmark of a useful measure. If that happens, we will most certainly throw out more than the bath water.

STEPHEN F. BUTLER

Inflexxion, Inc.

320 Needham Street

Suite 100

Newton

MA 02464-1594

USA

E-mail: sfbutler@inflexxion.com

DECLARATION OF INTEREST

Dr Butler is employed by Inflexxion, Inc., which has a commercial interest in a multimedia version of the Addiction Severity Index.

REFERENCES

- Budman, S. H., Portnoy, D. & Villapiano, A. (2003) How to get technological innovation used in behavioral healthcare: build it and they still might not come. *Psychotherapy: Theory, Research, Practice and Training*, **40**, 45–54.
- Butler, S. F., Budman, S. H., Goldman, R. J., Newman, F. L., Beckley, K. E., Trottier, D. & Cacciola, J. S. (2001) Initial validation of a computer-administered addiction severity index: the ASI-MV. *Psychology of Addictive Behaviors*, **15**, 4–12.
- Butler, S. F., Newman, F. L., Cacciola, J. S., Frank, A., Budman, S. H., McLellan, A. T., Ford, S., Blaine, J., Gastfriend, D., Moras, K. & Salloum, I. M. (1998) Predicting ASI interviewer severity ratings for a computer-administered Addiction Severity Index. *Psychological Assessment*, **10**, 399–407.
- Carise, D., McLellan, A. T. & Gifford, L. S. (2003) Developing a national addiction treatment information system. Available at: <http://www.densonline.org/DENSArt1update040101.pdf> (accessed online 2 December 2003).
- Geisinger, K. F. (1994) Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessments*, **6**, 304–312.
- Loevinger, J. (1954) The attenuation paradox in test theory. *Psychological Bulletin*, **51**, 493–504.
- Mäkelä, K. (2004) Studies of the reliability and validity of the Addiction Severity Index. *Addiction*, **99**, 398–410.
- Meehl, P. E. & Rosen, A. (1955) Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychology Bulletin*, **52**, 194–216.
- Nunnally, J. C. & Bernstein, I. H. (1994) *Psychometric Theory*, 3rd edn. New York: McGraw-Hill.
- Rogler, L. H. (1999) Methodological sources of cultural insensitivity in mental health research. *American Psychologist*, **54**, 424–433.
- Stewart, A. L., Hays, R. D. & Ware, J. E. (1992) Methods of constructing health measures. In: A. L. Stewart & J. E. Ware, eds. *Measuring Functioning and Well-Being: the Medical Outcomes Approach*, pp. 67–85. Durham, NC: Duke University Press.

- Streiner, D. L. (2003a) Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*, 99–103.
- Streiner, D. L. (2003b) Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, *80*, 217–222.
- Treatment Research Institute (TRI) (1990) Addiction Severity Index: Manual and Question by Question Guide. Available at: <http://www.densonline.org/qbyq.pdf> (accessed online 2 December 2003).

THE ASI: IS THE GLASS HALF EMPTY OR ALTOGETHER BRIMMING OVER?

Mäkelä (2004) is to be commended for this systematic review of the reliability and validity of the Addiction Severity Index (ASI), the more so because he largely follows the desiderata for reviews enumerated by Rehm (1999). The identification of reasonable study inclusion criteria is helpful, as is the paper's coherent argument and constructive suggestions. Very timely is the call to apply modern analytical and measurement techniques and common sense to the further refinement of the ASI, especially in comparative studies. After all, 'measurement remains a foundation of statistical inference' (Newman & Tejada 1999). A semistructured instrument such as the ASI, designed to serve both clinical appraisal and research goals, faces the challenge of optimizing for several purposes. The ASI represents a compromise in numerous ways articulated by its originators (e.g. McLellan *et al.* 1992). For example, in order to collect data with integrity and clinical meaning, the interviewer 'must be willing to repeat, paraphrase and probe until the patient understands the question and the answer is consistent with the intent of the question' (McLellan *et al.* 1992, p. 201). Such a strategy, while laudable from a clinical perspective, and possibly helping assure accurate reporting, may also leave answers excessively vulnerable to interviewer effects, even with careful training.

Mäkelä (2004) rightly notes that, given the multi-dimensional nature of the instrument, the ASI's authors, independent researchers and compendiums have tended to make overly sweeping categorical psychometric claims for the instrument in totality. For example, its composite scores are said to have 'generally excellent' reliability and 'demonstrated' concurrent, predictive, and discriminant validity (McLellan 2000). The emphasis on interviewer training requirements allows a naive reader to suppose that studies not achieving good results have involved flawed training: 'when interviewers were appropriately trained, these severity estimates were reliable and valid across a range of patients and interviewer types'

(McLellan *et al.* 1992; p. 202). Conversely, to their credit, the ASI's authors have encouraged other researchers to freely pursue improvements and have noted absence of reliability and validity studies in certain settings, especially ones where misrepresentation may be the rule rather than the exception, e.g. in probation/parole evaluations (McLellan *et al.* 1992). The extent to which routine assessment of treatment seekers can be correctly assumed to involve minimal bias deserves fuller study.

The contrast between such widely reflected enthusiasm and Mäkelä's more sober and discriminating appraisal raises the question of the merits of the practice, to which scale this author confesses succumbing himself (Greenfield & Attkisson 2004), of summarizing research on one's own instruments. On one hand, who knows the research more comprehensively? On the other hand, can a scale's originator be expected to be sufficiently objective in presenting both strengths and weaknesses? Of course, independent reviewers themselves may display partiality. One may quibble, for instance, at several of Mäkelä's pessimistic assessments, such as calling the associations for employment/support, drugs and legal status all 'alarmingly low' when in truth, for the latter two scales, only two and one, respectively, of seven studies shown for each in Table 4 resulted in a correlation lower than 0.50 (the criterion chosen). Similarly, in the critical Table 2, for employment status, eight of 11 studies result in alphas under the selected 0.70; but four values just missed, being 0.65 or above, with none below 0.50. For legal status, only three of the six failures fall below 0.65. None the less, taking a hard look at the broad psychometric claims made for the ASI remains essential, and it is most welcome that an emeritus master, re-emerging with such a tour-de-force review in the skeptical tradition, balances the accounting.

THOMAS K. GREENFIELD
 Alcohol Research Group
 2000 Hearst Avenue
 Suite 300
 Berkeley CA 94709
 USA
 E-mail: tgreenfield@arg.org

REFERENCES

- Greenfield, T. K. & Attkisson, C. C. (2004) The UCSF Client Satisfaction Scales. II. The Service Satisfaction Scale-30. In: Maruish, M., ed. *Psychological Testing: Treatment Planning and Outcome Assessment*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mäkelä, K. (2004) Studies of the reliability and validity of the Addiction Severity Index. *Addiction*, *99*, 398–410.
- McLellan, A. T. (2000) Addiction Severity Index (ASI). In: Rush, A. J., Pincus, H. A., First, M. B., Blacker, D., Endicott, J., Keith,