Drug and Alcohol Dependence 119 (2011) 229–234

Contents lists available at SciVerse ScienceDirect

CPDD News and Views

Society hompage: www.cpdd.vcu.edu

The promise of PROMIS® for addiction☆,☆☆

Thomas F. Hilton*

National Institute on Drug Abuse, National Institutes of Health, 6001 Executive Blvd, Rm 5197, Bethesda, MD 20892, United States

ARTICLE INFO

Article history:
Received 30 September 2011
Accepted 30 September 2011
Available online 1 November 2011

Keywords:
Measurement burden
Item Response Theory
Computerized adaptive testing
Classic test theory
Recovery monitoring

ABSTRACT

The field of addiction treatment has a measurement problem that pervades efforts to help patients achieve self-sustainable recovery. The impact of using older measurement technology has increased the measurement burden on both service providers and patients, while at the same time limiting the scope and frequency of measurement. The resulting burden can affect provider performance, patient access, and addiction recovery. This paper underscores the need for applying modern measurement theory techniques to reduce the measurement burden currently affecting most if not all major aspects of treatment and recovery. It is currently possible to obtain information more precisely, over a broad spectrum of recovery-oriented domains, faster and at lower cost than current measurement practices allow. However, a persistent research effort will be necessary to achieve that goal.

1. Introduction

It is no secret that the field of addiction treatment is suffering from the chronic relapsing problem of organizational inefficiency. Although researchers from leading institutions have identified remedies for many inefficient and ineffective business and therapeutic practices, the field has had difficulty adjusting to ever-changing requirements from outside organizations such as third-party payers and governmental agencies. Even when addiction treatment providers adopt efficiency and quality improvement programs like those advocated by the Network for the Improvement of Addiction Treatment, better known as NIATx; patient measures collected at intake and extending to self-sustainable recovery, remain technically inefficient. Current patient measures take too much time and resources to provide information that can be obtained across a broad scope of patient functioning, faster, cheaper, and with equal or greater precision.

The field's measurement problem pervades efforts to help patients achieve self-sustainable recovery. Waiting lists, which are almost a fixture among treatment providers, can be linked in part to the workload burden of inefficient measurement. Likewise, the high dropout rate among new admissions can be ascribed, in part, to the field's measurement burden. The associated response burden of lengthy intakes on patients, most of whom have diminished cognitive-emotional capacity, might be off-putting for many. The very act of monitoring progress, still a rare practice in addiction treatment, may be hamstrung by inefficient measurement. It costs both time and money to train, supervise, and administer lengthy measures leaving few resources to invest in monitoring progress. McLellan (2002), Dennis and Scott (2007), and others have argued, that monitoring recovery progress must begin during treatment, and extend long enough to attain self-sustainable recovery – a period that takes five or more years after treatment (McLellan et al., 2000; Betty Ford Institute Consensus Panel, 2007; DuPont et al., 2009; McKay et al., 2009). Results of a recent study by Crits-Christoph and Rotrosen (2011) showed that regular progress monitoring indeed can improve counselors' success in keeping patients on the path to recovery. Such a practice is unlikely to spread until the field has better tools to monitor outcomes with efficiency and precision.

2. Patient self-reports

The vast majority of measures used in both addiction research and treatment rely upon patient self reports (PSR's). Some PSRs' involve cataloguing clinically-relevant characteristics and assessments of patient symptoms (APS's). Others involve patient reported outcomes (PRO's) that provide information about how treatment is affecting progress toward recovery. APS's are used clinically to diagnose and select a course of treatment. They are usually obtained at or near the time of admission, and often

☆ This material is not peer-reviewed by the Journal, but is reviewed prior to publication by the members of the CPDD Publications Committee and invited members of the College. News and Views is edited by the Chair of the CPDD Publications Committee: Gregory M. Miller, Ph.D., Harvard Medical School, New England Primate Research Center, Pine Hill Drive, Southborough, MA 01772, USA.

☆☆ The views and opinions expressed in this report are those of the author and should not be construed to represent the views of NIDA or any of the sponsoring organizational, agencies, or the U.S. Government.

* Tel.: +1 301 435 0808; fax: +1 301 443 6815.
E-mail address: tom.hilton@nih.gov

0376-8716/$ – see front matter
doi:10.1016/j.drugalcdep.2011.09.031

include brief screeners – typically single gateway questions that justify ruling out certain diagnoses or that may lead to deeper probes depending on the answer. In research, APS's are often used to quantify certain patient characteristics that might differentially affect treatment outcomes. PRO's, on the other hand, are used clinically to monitor the effects of interventions when the course of treatment is expected to extend for a significant period of time. In research, PRO's are used to quantify change over time, and may be used to make differential distinctions in clinical trials comparing different therapeutic or other interventions (US Food and Drug Administration, 2009).

The reason for making what may seem at first blush to be an arbitrary distinction among PSR's is that the intended use of a PSR, assessment or outcomes monitoring, has important implications for calibration and validation. Many, APS's (especially brief screeners) are idiosyncratic to individual patients, are infrequently repeated, and thus were not designed or validated for use in monitoring change nor for enabling comparisons between patient outcomes, i.e., comparative effectiveness research. Demonstrating sensitivity to change over time is rarely part of APS validation design, because assessments are mainly used for diagnosis, treatment planning, and trait classification. The calibration of APS scores typically emphasizes precision on the symptomatic side of the distribution. This means that most APS's tend to be less precise in detecting gradations in patient wellness than illness, and thus less precise for quantifying recovery progress. The application of PRO's is quite different. Although some PRO's may be valid for assessment purposes, they are intended primarily for use in monitoring the effect that treatment is having over time (Patrick et al., 2008). Thus, the validation goals of PRO's include: (a) brevity – low patient response burden, (b) precision – over a broad range of diseases, symptoms, functionality, and levels of severity, (c) reliability – over an extended period of time – if not the remaining life of the patient, and (d) validity – for the study of trends and comparative treatment effectiveness. Clearly, such attributes are desirable for all types of PSR's. Unfortunately, most PSR's used in general healthcare today fail to exhibit such ideal psychometric characteristics. However, I will remark on a few notable exceptions.

The remainder of this paper will focus mainly on the need to reduce the measurement burden of current PSR's used in addiction treatment and recovery. Issues of precision, reliability, and validity will be addressed, but within the context of informing readers that a change in measurement practices is not only possible, but necessary.

3. Measurement burden

The PSR measurement burden on patients entering and receiving addiction treatment services, as well as those providing those services, is not trivial (Ford et al., 2007), and has been linked to clinic performance (Carise et al., 2009; McCarty et al., 2007a; Rupert and Morgan, 2005). The PSR measurement burden also may affect treatment engagement and ultimately the pace of recovery from addiction.

3.1. Patient burden

Lengthy patient interviews can be tiring for patients suffering from any illness, injury, or disability. In the case of addiction, the added cognitive-emotional dysfunction due to brain toxicity accentuates the assessment burden (Bates et al., 2006; Meek et al., 1989; Vik et al., 2004), while at the same time increasing the likelihood of cognitive distortions that can affect reliability and validity (Bendig, 1955; Thorndike et al., 1951). Often seeking treatment under some degree of duress, addiction patients may experience ambivalence about enrollment. Intake is not the best time to be providing reasons to drop out by subjecting patients to lengthy, sometimes redundant, questioning under conditions of heightened impaired judgment (see Katz et al., 2005).

3.2. Provider burden

Lengthy PSR protocols also have an impact on individual service providers (McLellan and Meyers, 2004). Not only can lengthy interviews reduce time available for therapeutic activities, but length too can reduce reliability and validity due to fatigue and interviewer drift (Fureman et al., 1994).

An additional concern related to the measurement burden in addiction is that the time and expense of lengthy assessments consumes organizational resources. For example, one study reported that over half of the total cost of addiction treatment services (inpatient or outpatient) paid for intake activities (Anderson et al., 1998) – most of which involve PSR's. Another budgetary aspect of measurement is how it affects training costs. Most assessments in current use require formal training and supervision to administer reliably and validly, and/or to score. Given the high annual turnover rate among the nation's counselors (shown to be over 18% nationally; Knudsen et al., 2003; over 33% in some settings; Eby et al., 2010), reducing the cost of training new personnel can have a significant effect on operating budgets. Even merely computerizing PSR measures can greatly reduce the need for training to administer APS's and PRO's (e.g., Butler et al., 2001).

Finally, improved measurement efficiency enables enhanced quality of care. By freeing up time and other resources, it is possible to measure and monitor topics commonly asked of general medical patients such as fatigue, pain, physical functionality, and sleep. These and other more general-medical symptoms have been linked to addiction and relapse (e.g., Ramo and Brown, 2008; Morgan et al., 2003; See reviews by Fishbain et al., 2008; McKay and Weiss, 2001). Though they might not be manifest at intake, it is not uncommon for new symptoms to emerge during the course of recovery (Dennis et al., 2007). Thus, PRO's need to include a broad range of domains to detect emergent symptoms not identified during intake.

3.3. Treatment engagement

The measurement burden problem in addiction treatment almost certainly plays a role in dropping out of treatment and relapse. Addiction treatment studies have repeatedly indicated that most dropouts occur within days of intake assessments (Hoffman et al., 2011; De Leon and Schwartz, 1984; Pulford and Wheeler, 2007). Indeed, the Network for Improving Addiction Treatment (NIATx) has demonstrated that streamlined intake procedures can significantly increase patient access and decrease treatment dropouts (McCarty et al., 2007b). Studies have suggested that burdensome intake procedures can be off-putting to experience (Scott and Foss, 1999), and there is some evidence that less intensive procedures can enhance treatment retention (Bell et al., 1994; Woody et al., 1975). Lingering cognitive-emotional impairment due to brain toxicity is most acute in the months following detoxification when most patients are entering treatment. This makes the concentration necessary to undergo hours of interviews and assessments difficult and unpleasant for patients. It can raise doubts about the value of treatment participation, especially if assessments appear repetitive or pointless.

3.4. Recovery

McLellan has underscored that treating addiction as a chronic disease

requires regular outcomes monitoring during formal treatment and continuing into recovery support aftercare (McLellan, 2002). Such regular measurement is rare, doubtless stymied by a seemingly ever-increasing measurement burden (McLellan et al., 2005). Without the PRO monitoring called for by McLellan, counselors and case managers are left with little more than an intake assessment, loose recollections of numerous clinical impressions formed across therapy sessions, and various bits of observational information provided by colleagues and family to guide their therapeutic decision-making. In addition, Dennis et al. (2006) point out that the time and costs of staff training plus the time to administer lengthy assessments often exceeds available resources. Resource limitations may explain why treatment providers often limit their measurement resources to focus on social–emotional adjustment and substance use behaviors. Routine, systematic, precise, collection of information about patient pain, physical functioning, and sleep/wake problems – each of which can play a role in addiction and recovery – is not common.

4. Modern measurement theory

The addiction field's approach to assessment and monitoring PRO's is burdensome and imprecise, yet technology exists to improve the situation. The vast majority of measures used in addiction treatment and recovery were developed based on what is now called classic test theory (CTT; see Lord and Novick, 1968). CTT measures are constructed of items sampled from the domain of factors considered relevant to the topic of interest. The aggregate of the responses to those items becomes the score for that domain/topic/measure. Under CTT, measurement error is considered to be averaged out, i.e., positive and negative errors cancelling each other out (Hambleton and Jones, 1993). Reliability and validity analyses are conducted relative to the effect items have on the total score. Under CTT, score precision requires that all the items be addressed by respondents. If fewer items are preferred for a measure, revalidation of the shorter measure is required, and almost certainly, there will be a loss in precision as dropping items will reduce the scope of the domain addressed. Alternatively, if greater precision is desired, developers broaden the sample of domain-related items, which in nearly all cases, increases the number of items.

In contrast to CTT, modern measurement theory (MMT) addresses measurement at the item level. MMT assumes that each item is a valid indicator of a respondent's position on a unidimensional domain, and the ability to correctly answer or otherwise endorse a self-report item as a true reflection of one's condition allows for arranging respondents on an endorsement difficulty continuum. Those endorsing or correctly answering the most difficult items reflect greater presence of a trait or symptom (e.g., mathematical reasoning skill, depression, pain). Under MMT, each item is assessed for validity and reliability, as well as response characteristics. Items indicating low precision in initial field tests are replaced with items showing greater precision. Unlike CTT measurement, under MMT, the effect on precision of adding or deleting any item to the derived score is always known. This fact facilitates replacing or improving problematic items such as those that might be moderated by extraneous factors such as gender, reading level, or race, which usually reduces the error associated with MMT scores when compared to CTT scores (Bjorner et al., 2007).

Studies have shown that CTT and MMT measures of the same trait/domain often produce scores of similar reliability and validity when the same number of items is used and measures are developed with comparable rigor (e.g., Fan, 1998; Hambleton and Jones, 1993; Singh, 2004). Where CTT becomes problematic is when the goal is to minimize the number of items used without sacrificing validity, as is the goal of computerized adaptive test (CAT) measures. MMT makes it possible to develop brief, precise, reliable, and valid CAT measures.

Modern Measurement Theory (MMT) emerged in the 1950s with the advent of articles on Item Response Theory (IRT) by Fredrick Lord (Lord, 1950, 1980) and Warren Torgerson (Torgerson, 1958). During the 1970s, practical applications of MMT to CAT were being applied to educational and military recruit aptitude-testing (Weiss and Kingsbury, 1984; Weisen and Siegel, 1976) as a way to reduce the time to assess tens of thousands of people every year. Today, applications of CAT technology in healthcare settings are rapidly expanding. The NIH Patient Reported Outcomes Measurement Information System (PROMIS®) is one example of a general health PRO battery that uses CAT technology to precisely assess a dozen diverse general health domains in under fifteen minutes that would otherwise take more than an hour (Khanna and Hays, 2011). More NIH PROMIS® domains are being added monthly.

Unfortunately, not all assessment items meet the requirements for CAT technology as some APS items fail to meet IRT assumptions of unidimensionality or exhibit minimum response characteristics to assure precision (Riley et al., 2007). In some cases, there are work-arounds. In other cases, CAT measurement is unworkable, and it is necessary to rely upon brief CTT scales. However, even measures inappropriate for CAT applications can often be improved upon by computerization. Computers can easily apply Guttman-type inquiry trees that can shorten assessments by skipping unnecessary items or scales. For example, if a patient indicates never having used cocaine, there is no need to ask follow-up questions about history and frequency of cocaine use. Although a few addiction researchers have undertaken the challenge to develop both computerized (e.g., ASI-MV; Butler et al., 2001) and CAT measures (e.g., D-CAT; Fliege et al., 2005; GAIN-SPS; Riley et al., 2007) to replace popular measures used in addiction, the field has an urgent need for upgrading its measurement inventory.

A relatively new evolving area of MMT is Multidimensional IRT (MIRT; Ackerman, 1994; Walker and Beretvas, 2003). MIRT not only allows for using the same item to measure more than one dimension (thereby further reducing response burden), but also enables addressing constructs with more than one dimension, which is frequently the case for mental health assessments (Segall, 1996; Gardner et al., 2002). MIRT is now enabling the development of CAT assessment measures for domains used to diagnose and plan treatment for comorbid addiction patients.

4.1. CAT properties

Following Louis Guttman's (1968) seminal work, modern CAT technology uses computers to search among items in a calibrated bank of PRO items representing a single unidimensional latent trait/domain such as fatigue. Ideally, each CAT item is culled from a near-exhaustive legacy item bank derived from the published literature (Klem et al., 2009). Though not essential to building item banks, starting with an exhaustive preliminary bank helps ensure that popular legacy items are included in the final CAT banks as long as individual items meet precision requirements. Within the healthcare field, focus groups with researchers, patients, and care providers help to identify any gaps, improve clarity, and introduce experimental complementary items.

CAT items are ordered by symptom severity reflected in item endorsement difficulty. Each item found to validly indicate a patient's location on the continuum of a PRO dimension ranging from well to highly symptomatic is calibrated individually. Calibration aims to ensure common interpretation by respondents, acceptable discrimination across response options, sensitivity to changes in condition across a broad spectrum of intensity, invariance across respondent characteristics and health conditions, and validity with which responses relate to actual health status (Reeve et al., 2007). This sort of precision is very difficult to achieve for CTT measures because they are not calibrated at the item level (Magno, 2009; McDonald and Paunonen, 2002; Singh, 2004; Adedoyin et al., 2008).

4.2. How CAT works

Fixed-item measures developed under CTT estimate severity using total symptom counts. By contrast, CAT re-estimates symptom severity after each response and determines if a significant change from the prior responses has occurred. If not, it stops – usually after asking fewer items than fixed-item scales developed under CTT. CAT's operate much the same as hearing tests, in which one can no longer press a button indicating their ability to hear increasingly difficult-to-detect sounds. Scores relative to the population norm for each dimension indicate the presence or absence of pathology. Developing valid CAT measures is time-consuming and costly, however, the final product is an enormous gain in speed and precision (Riley et al., 2007).

4.3. CAT in general healthcare monitoring

To date, the National Institutes of Health (NIH) has invested approximately $100 million dollars in modernizing PRO assessment. As mentioned above, PROMIS® represents the largest and most comprehensive attempt to replace the current "PRO Tower of Babel" with tools that are far faster, better, and cheaper to use than CTT-derived measures in which a response to all items is required to derive a score (Cella, 2010, 2011). The NIH Toolbox for Assessing Neurological and Behavioral Function (Toolbox) and the Neurological Quality of Life (Neuro-QOL) batteries also represent companion efforts by NIH to modernize PRO's which will eventually be considered for PROMIS® inclusion. Both Toolbox and Neuro-QOL augment PROMIS® domains, and are designed similarly to obtain fast, precise PRO data from patients suffering from neurological disorders.[1]

General medicine colleagues might wonder why patient outcomes commonly monitored in most other diseases are overshadowed by measures of drug craving, criminality, and mental illness. A large percentage of addiction patients can be suffering from brain trauma (Walker et al., 2007), chronic lifestyle-related diseases such as heart disease and diabetes (Hser et al., 2001; Asgary et al., 2008), as well as chronic pain resulting from trauma and/or surgeries. Moreover, despite research linking addiction to such symptoms as physical pain (Fishbain et al., 2008), general depression and anger (e.g., Rush et al., 2008), as well as sleep disorders (e.g., Friedmann et al., 2003), these topics are rarely well-assessed during intake processing, and are even more unlikely to be monitored during the course of addiction treatment and recovery. Thus, the field of addiction knows relatively little about patient physical health compared to what it knows about patient drug use behaviors, criminality, and mental health. This systemic blindness will persist unless the field begins to monitor PRO's on a broader scope of domains than is current practice. The new, freely available, NIH PROMIS® measures could be added tomorrow were it not for the legacy of burdensome measures still in common use within the addiction field.

4.4. CAT research in addiction APS's and PRO's

Several projects are in the field that are developing and validating addiction-relevant APS's and PRO's. NIDA has awarded grants to develop two new PRO item banks. The first grant is to the RAND Corporation (PI Maria Orlando-Edelen) to develop a PRO item bank addressing smoking behaviors. The second is to UCLA (PI Li Cai) to explore development of new CAT item banks addressing changes in addiction recovery status. Both projects will be compatible with PROMIS®. NIDA also has awarded grants to begin modernizing addiction APS's. Inflexxion (PI Stephen Butler) is currently exploring ways to apply CAT technology to its computerized web-based version of the popular Addiction Severity Index, the ASI Multi-media Version (ASI-MV); and Chestnut Health Systems (PI: Barth Riley) is exploring ways to apply MMT to reduce the response burden associated with the Global Appraisal of Individual Needs (GAIN). NIDA's grants to UCLA and Chestnut Health Systems both include MIRT research applications to addiction assessments. Finally, a PROMIS® item bank to address alcohol use has recently been completed at the University of Pittsburgh (PI Paul Pilkonis), and is ready for validation research. The alcohol bank calibration research was conducted using NIDA Clinical Trials Network patients, who also completed the PROMIS® CAT general health battery of item banks.

Why adopt the new NIH PRO measures? Monitoring PRO's, especially during recovery, enhances the likelihood of detecting potential relapse triggers or early warning signs. An increase in sleep-related PROs or physical pain could alert service providers to a possible relapse risk early enough to intervene before substance use is resumed. In the past, given the resource burden, suggesting the addition of a dozen general health measures to an intake protocol would be met with disbelief. However, as it normally takes less than 15 minutes for PROMIS® CAT to achieve that goal, it is not unreasonable to consider expanded use of PROs to help advance treatment and recovery. As less burdensome measures enter the field, especially computerized ones, the field can look forward to greater access to treatment that is of higher quality and better suited to promote addiction recovery.

Concerns about replacing legacy data with new measures are reasonable, but in many cases unfounded because valid legacy items are in PROMIS® measures already. That fact helps to ensure that new PROMIS® CAT measures are tau equivalent with popular legacy measures of the same domain. Because of relatively better precision and sensitivity ranges common in PROMIS® measures, they should offer superior monitoring utility when following patients progress over time.

PROMIS® already has the capability of uploading results directly to patients' electronic health records (EHR's) within seconds of completing the battery. PROMIS® was designed to work in EHR environments using SNOMED (Systematized Nomenclature of Medicine) and LOINC (Logical Observation Identifiers Names and Codes). Thus, it is already possible to conduct healthcare systems monitoring and research using PROMIS® data. Ideally, all computerized addiction measures would someday have EHR capability even if not fully CAT capable. However, that is unlikely to happen

[1] More information about Neuro-QOL and Toolbox is available at the PROMIS® Assessment Center website where readers can build and use PROMIS® measures and obtain more psychometric background on PROMIS: www.assessmentcent.net.

unless the field starts now to invest time and resources to the necessary research and development.

While NIDA and other NIH institutes and centers continue current MMT initiatives, the entire field can and should get involved. Both researchers and service providers should set a new standard for any measure they use. If a computerized version is available – explore its relative advantages. If a CAT version is available, I hope this article has made a convincing case for adoption over paper and pencil measures developed under CTT. Finally, because PROMIS® is freely available for implementation today,[2] and will soon include new item banks for measuring alcohol use and smoking, it should be considered for inclusion as the centerpiece of measurement modernization initiatives in both research and practice. PROMIS®, and other modern measures like it, offer the field new tools for extending recovery for more patients for a long enough period to reach self-sustained sobriety.

References

Ackerman, T., 1994. Using multidimensional Item Response Theory to understand what items and tests are measuring. Appl. Meas. Educ. 7, 255–278.

Adedoyin, O.O., Nenty, H.J., Chilisa, B., 2008. Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. Educ. Res. Rev. 3 (2), 83–93.

Anderson, D.W., Bowland, B.J., Cartwright, W.S., Bassin, G., 1998. Service-level costing of drug abuse treatment. J. Subst. Abuse Treat. 15 (3), 201–211.

Asgary, S., Sarrafzadegan, N., Naderi, G.-A., Rozbehani, R., 2008. Effect of opium addiction on new and traditional cardiovascular risk factors: do duration of addiction and route of administration matter? Lipids Health Dis. 7 (42), Available from: http://www.lipidworld.com/content/7/1/42 (accessed 05.05.11).

Bates, M.E., Pawlak, A.P., Tonigan, J.S., Buckman, J.F., 2006. Cognitive impairment influences drinking outcome by altering therapeutic mechanisms of change. Psychol. Addict. Behav. 20, 241–253.

Bell, J., Caplehorn, J.R.M., McNeil, D.R., 1994. The effect of intake procedures on performance in methadone maintenance. Addiction 89, 463–471.

Bendig, A.W., 1955. Rater reliability and "Judgmental Fatigue". J. Appl. Psychol. 39 (6), 451–454.

Betty Ford Institute Consensus Panel, 2007. What is recovery? A working definition from the Betty Ford Institute. J. Subst. Abuse Treat. 33, 221–228.

Bjorner, J.B., Chang, C.-H., Thissen, D., Reeve, B.B., 2007. Developing tailored instruments: item banking and computerized adaptive assessment. Qual. Life Res. 16, 95–108.

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. Psychol. Addict. Behav. 15, 4–12.

Carise, D., Love, M., Zur, J., McClellan, A.T., Kemp, J., 2009. Results of a statewide evaluation of "paperwork burden" in addiction treatment. J. Subst. Abuse Treat. 37, 101–109.

Cella, D., 2010. New approaches to measuring important aspects of cancer survivorship: PROMIS and the Cancer PROMIS Supplement (CaPS). In: Presentation at the 2010 Biennial Conference on Cancer Survivorship Research, Washington, DC, June 18.

Cella, D., 2011. PROMIS® Program Overview. In: Presentation to the National Institutes of Health, January 11, 2011, National Institutes of Health Bldg 31, Bethesda, MD.

Crits-Christoph, P., Rotrosen, J., 2011. Personal communication regarding completed research not yet in peer review showing that counselors provided patient-level feedback monitoring (when compared to those who did not) were significantly more likely to help sustain progress toward recovery.

De Leon, G., Schwartz, S., 1984. Therapeutic Communities: what are the retention rates? Am. J. Drug Alcohol Abuse 10 (2), 267–284.

Dennis, M.L., Scott, C.K., 2007. Managing addiction as a chronic condition. Addict. Sci. Clin. Pract. 4 (1), 45–55.

Dennis, M.L., Chan, Y.-F., Funk, R.R., 2006. Development and validation of the GAIN Short Screener (GSS) for Internalizing Externalizing, and Substance Use Disorders and Crime/Violence Problems among Adolescents and Adults. Am. J. Addict. 15, 80–91.

Dennis, M.L., Foss, M.A., Scott, C.K., 2007. An eight-year perspective on the relationship between the duration of abstinence and other aspects of recovery. Eval. Rev. 31 (6), 585–612.

DuPont, R.L., McLellan, A.T., White, W.L., Merlo, L.J., Gold, M.S., 2009. Setting the standard for recovery: Physicians' Health Programs. J. Subst. Abuse Treat. 36, 159–171.

Eby, L.T., Burk, H., Maher, C.P., 2010. How serious of a problem is staff turnover in substance abuse treatment? A longitudinal study of actual turnover. J. Subst. Abuse Treat. 39, 264–271.

Fan, X., 1998. Item Response Theory and classical test theory: an empirical comparison of their item/person statistics. Educ. Psychol. Meas. 58 (3), 357–381.

Fishbain, D.A., Cole, B., Kewis, J., Rosomoff, H.L., Rosomoff, R.S., 2008. What percentage of chronic nonmalignant pain patients exposed to chronic opioid analgesic therapy develop abuse/addiction and/or aberrant drug-related behaviors? A structured evidence-based review. Pain Med. 9 (4), 444–459.

Fliege, H., Becker, J., Walter, O.B., Bjorner, J.B., Klapp, B.F., Rose, M., 2005. Development of a computer-adaptive test for depression (D-CAT). Qual. Life Res. 14, 227–2291.

Ford, J.H., Green, C.A., Hoffman, K.A., Wisdom, J.P., Riley, K.J., Bergman, L., Molfenter, T., 2007. Process improvement needs in substance abuse treatment: admissions walk-through results. J. Subst. Abuse Treat. 33 (4), 379–389.

Friedmann, P.D., Herman, D.S., Freedman, S., Lemon, S.C., Ramsey, S., Stein, M.D., 2003. Treatment of sleep disturbance in alcohol recovery: a national survey of addiction medicine physicians. J. Addict. Dis. 22 (2), 91–103.

Fureman, I., McLellan, A.T., Alterman, A., 1994. Training for and maintaining interviewer consistency with the ASI. J. Subst. Abuse Treat. 11 (3), 223–237.

Gardner, W., Kelleher, K.J., Pajer, K.A., 2002. Multidimensional adaptive testing for mental health problems in primary care. Med. Care 40, 812–823.

Guttman, L., 1968. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika 33 (4), 469–506.

Hambleton, R.K., Jones, R.W., 1993. Comparison of classical test theory and Item Response Theory and their applications to test development. Educ. Meas. Issues Pract. 12, 38–47.

Hoffman, K.A., Ford, J.H., Tillotson, C.J., Choi, D., McCarty, D., 2011. Days to treatment and early retention among patients in treatment for alcohol and drug disorders. Addict. Behav. 36, 643–647.

Hser, Y.-I., Hoffman, V., Grella, C.E., Anglin, M.D., 2001. A 33-year follow-up of narcotics addicts. Arch. Gen. Psychiatry 58, 503–508.

Katz, E.C., King, S.D., Schwartz, R.P., Weintraub, E., Barksdale, W., Robinson, R., et al., 2005. Cognitive ability as a factor in engagement in drug abuse treatment. Am. J. Drug Alcohol Abuse 31, 359–369.

Khanna, D., Hays, R., 2011. Personal communication with Dinesh Khanna and Ron Hays from UCLA on April 26, 2011. In a study currently in review, 143 patients suffering from scleroderma had median response times for 11 PROMIS® general health domains of 9 minutes with a mean of 11.9 minutes.

Klem, M., Saghafi, E., Abromitis, R., Stover, A., Dew, M.A., Pilkonis, P., 2009. Building PROMIS item banks: librarians as co-investigators. Qual. Life Res. 18, 881–888.

Knudsen, H.K., Johnson, J.A., Roman, P.M., 2003. Retaining counseling staff at substance abuse treatment centers: effects of management practices. J. Subst. Abuse Treat. 24, 129–135.

Lord, F.M., 1950. A theory of test scores. Psychol. Monogr. 7.

Lord, F.M., 1980. Application of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum, Hillsdale, NJ.

Lord, F.M., Novick, M.R., 1968. Statistical Theories of Mental Test Scores. Addison-Wesley, Reading, MA.

Magno, C., 2009. Demonstrating the difference between Classical Test Theory and Item Response Theory using derived test data. Int. J. Educ. Psychol. Assess. 1 (1), 1–11.

McCarty, D., Gustafson, D.H., Wisdom, J.P., Ford, J., Dongseok, C., Molfenter, T., Cappocia, V., Cotter, F., 2007a. The Network for the Improvement of Addiction Treatment (NIATx): enhancing access and retention. Drug Alcohol Depend. 88, 138–145.

McCarty, D., Gustafson, D.H., Wisdom, H.P., Ford, J., Choi, D., Molfenter, T., Capoccia, V., Cotter, F., 2007b. The Network for the Improvement of Addiction Treatment (NIATx): enhancing access and retention. Drug Alcohol Depend. 88 (2–3), 138–145.

McDonald, P., Paunonen, S.V., 2002. A Monte Carlo comparison of item and person statistics based on Item Response Theory versus Classical Test Theory. Educ. Psychol. Meas. 62 (6), 921–943.

McKay, J.R., Carise, D., Dennis, M.L., DuPont, R., Humphreys, K., Kemp, J., Reynolds, D., White, W., Armstrong, R., Chalk, M., Haberle, B., McLellan, T., O'Connor, G., Pakull, B., Schwarzlose, J., 2009. Extending the benefits of addiction treatment: practical strategies for continuing care and recovery. J. Subst. Abuse Treat. 36, 127–130.

McKay, J.R., Weiss, R.V., 2001. A review of temporal effects and outcome predictors in substance abuse treatment studies with long-term follow-ups. Eval. Rev. 25 (2), 113–161.

McLellan, A.T., Meyers, K., 2004. Contemporary addiction treatment: a review of systems problems for adults and adolescents. Biol. Psychiatry 56, 764–770.

McLellan, A.T., 2002. Have we evaluated addiction treatment correctly? Implications from a chronic care perspective. Addiction 97, 249–252.

McLellan, A.T., Lewis, D.C., O'Brien, C.P., Kleber, H.D., 2000. Drug dependence, a chronic medical illness. JAMA 284 (13), 1689–1695.

McLellan, A.T., McKay, J.R., Forman, R., Cacciola, J., Kemp, J., 2005. Reconsidering the evaluation of addiction treatment: from retrospective follow-up to concurrent recovery monitoring. Addiction 100, 447–458.

Meek, P.S., Clark, H.W., Solana, V.L., 1989. Neurocognitive impairment: the unrecognized component of dual diagnosis in substance abuse treatment. J. Psychoactive Drugs 21, 153–160.

Morgan, T.J., Morgenstern, J., Blanchard, K.A., Labouvie, E., Bux, D.A., 2003. Health-related quality of life for adults participating in outpatient substance abuse treatment. Am. J. Addict. Psychiat. 12, 198–210.

[2] Information regarding PROMIS® and its implementation can be found at www.NIHPROMIS.org.

Patrick, D.L., Guyatt, G.H., Acquadro, C., 2008. Patient-reported outcomes. In: Higgins, J., Green, S. (Eds.), The Cochrane Library. , 7th ed. Wiley, Chichester, England (Chapter 17).

Pulford, J., Wheeler, A., 2007. Documenting client attendance norms: raw data and implications for treatment practice. J. Subst. Use 12 (2), 95–102.

Ramo, D.E., Brown, S.A., 2008. Classes of substance abuse relapse situations: a comparison of adolescents and adults. Psychol. Addict. Behav. 22 (3), 372–379.

Reeve, B.B., Hays, R.D., Bjorner, J.B., Cook, K.F., Crane, P.K., Teresi, J.A., Thissen, D., Revicki, D.A., Weiss, D.J., Hambleton, R.K., Liu, H., Gershon, R., Reise, S.P., Lai, J.-S., Cella, D., et al., 2007. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient Reported Outcomes Measurement Information System (PROMIS. Med. Care 45 (Suppl. 1 (5)), S22–S31.

Riley, B.B., Conrad, K.J., Bezruczko, N., Dennis, M.L., 2007. Relative precision, efficiency and construct validity of different starting and stopping rules for a computerized adaptive test: The GAIN Substance Problem Scale. J. Appl. Meas. 8 (1), 48–64.

Rupert, P., Morgan, D., 2005. Work setting and burnout among professional psychologists. Prof. Psychol. Res. Pract. 36, 544–550.

Rush, B.R., Dennis, M.L., Scott, C.K., Castel, S., Funk, R.R., 2008. The interaction of co-occurring mental disorders and recovery management checkups on substance abuse treatment participation and recovery. Eval. Rev. 32 (1), 7–38.

Scott, C., Foss, M., 1999. The impact of centralized intake on treatment access and intake satisfaction. In: Paper presentation at the August American Psychological Association meeting, Boston.

Segall, D.O., 1996. Multidimensional adaptive testing. Psychometrika 61, 331–354.

Singh, J., 2004. Tackling measurement problems with Item Response Theory: principles, characteristics and assessment, with an illustrative example. J. Bus Res. 57 (2), 184–208.

Thorndike, R.L., with Cronbach, L.J., Cureton, E.E., Kelley, T.L., Kurtz, A.K., Richardson, M.W., Thurstone, L.L., 1951. Reliability. In: Lindquist, E.F. (Ed.), Educational Measurement. American Council on Education, Washington, DC, pp. 560–620.

Torgerson, W.S., 1958. Theory & Methods of Scaling. Wiley, New York.

US Food and Drug Administration, 2009. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims, www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf (accessed 01.02.11).

Vik, P., Cellucci, T., Jarchow, A., Hedt, J., 2004. Cognitive impairment in substance abuse. Psychiatr. Clin. N. Am. 27, 97–109.

Walker, C.M., Beretvas, S.N., 2003. Comparing multidimensional and unidimensional proficiency classifications: multidimensional IRT as a diagnostic aid. J. Educ. Meas. 40, 255–275.

Walker, R., Cole, J.E., Lohan, T.K., Corrigan, J.D., 2007. Screening substance abuse treatment clients for traumatic brain injury: prevalence and characteristics. J. Head Trauma Rehabil. 22 (6), 360–367.

Weisen, J.P., Siegel, A.I., 1976. Psychometric Characteristics of the Armed Services Vocational Aptitude Battery. ASVAB. Applied Psychological Services, Oxford, England.

Weiss, D.J., Kingsbury, G., 1984. Application of computerized adaptive testing to educational problems. J. Educ. Meas. 21 (4), 361–375.

Woody, G., O'Hare, K., Mintz, J., O'Brien, C., 1975. Rapid intake: a method for increasing retention rate of heroin addicts seeking methadone treatment. Compr. Psychiatry 16 (2), 165–169.