Published in final edited form as:

Drug Alcohol Depend. 2012 May 1; 122(3): 236–240. doi:10.1016/j.drugalcdep.2011.10.006.

The Physicians' Competence in Substance Abuse Test (P-CSAT): A Multidimensional Educational Measurement Tool for Substance Abuse Training Programs*

John M. Harris Jr.* and Huaping Sun**

*Medical Directions, Inc. and the Office of Continuing Medical Education, University of Arizona College of Medicine, both in Tucson, AZ 85712

**Department of Counseling, Educational Psychology and Research, the University of Memphis, Memphis, TN 38152

Abstract

Background—Efforts to enhance physician substance abuse (SA) skills lack tools to assess skills and training effectiveness. We sought to develop an easily administered survey to assess SA fact-based skills and clinical decision-making.

Methods—We prepared 60 fact-based items dealing with SA knowledge, attitudes, and behaviors (KAB) and 53 script concordance test (SCT) items assessing SA decision-making. We used expert review and standard psychometric criteria to eliminate discordant or non-contributory survey items. We tested 92 draft items in 117 physicians, including 13 with additional SA training (trained), and 17 recognized SA experts. We assessed final survey internal consistency with Cronbach's alpha and differences in scores between experts, trained physicians, and physicians without SA training (novices) with the Kruskal-Wallis test.

Results—Following refinement, the draft survey was reduced to 30 KAB and 33 SCT items. Alpha was 0.901 for the final 63-item survey and 0.887 and 0.797 for the KAB and SCT subscales, respectively. Novices, trained physicians, and experts scored means of 196, 213, and 261 respectively out of 315 possible points on the final survey. The KAB and SCT subscale results showed similar patterns. Score differences for the overall survey and its subscales were highly significant ($P < 0.001$).

Conclusions—This survey, which we have named the Physicians' Competence in Substance Abuse Test (P-CSAT) and placed in the public domain, meets baseline criteria for reliability and

*The P-CSAT items, item types, and scoring methods can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

© 2011 Elsevier Ireland Ltd. All rights reserved

Corresponding Author: John M. Harris Jr. Medical Directions, Inc. 6101 E Grant Rd., Tucson, AZ 85712 Phone: 520-722-1970 Fax: 520-722-2670 sharris@md-inc.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Contributors. Both authors designed the study. Dr. Harris wrote the protocol and managed the literature searches and summaries of previous related work. Dr. Sun undertook the statistical analysis. Dr. Harris wrote the first draft of the manuscript. Both authors contributed to and have approved the final manuscript.

Conflict of Interest. Dr. Harris is the president of Medical Directions, Inc., which owns the CME Website (www.vlh.com) used to recruit study subjects. Medical Directions, Inc. is also the owner of the P-CSAT. Dr. Sun has nothing to disclose.

validity. Future studies should determine the extent to which the P-CSAT provides consistent results in other practitioner populations and responds to SA educational efforts.

Keywords

Medical Education; Substance Abuse Training; Script Concordance Test

1. Introduction

The US national drug control strategy calls for early screening and intervention by health care professionals, but notes that, “Unfortunately, health care providers often do not have the training or resources to identify and treat patients' underlying issues of substance abuse.” (Office of National Drug Control Policy, 2011). There have been federally-funded efforts to remedy training deficits, particularly support for training in screening, brief intervention, and referral to treatment (SBIRT), but such funding has been reduced or eliminated in current NIH budgets. A recent study of SBIRT in a real-world rural setting found positive outcomes for both alcohol and drug abuse (Gryczynski et al., 2011). Thus, there is a need for wider training of health professionals to recognize and manage substance abuse (SA) and an increasing acceptance of SBIRT as a useful SA intervention, but much of this training will need to occur at local levels.

Unfortunately, there is little information on how to measure the success of drug abuse training programs that incorporate evolving SBIRT principles. This lack of useful measurement tools impedes efforts to develop successful and replicable provider training programs. As is true for most health professional education interventions, published reports of SA education programs focus mainly on program features and only briefly describe program measurement and assessment techniques. SA educational measurement tools are not typically evaluated for reliability or validity beyond face validity and have included (locally) standardized patients, student feedback, knowledge recall, and observer assessment (Bernstein et al., 2007; Parish et al., 2006; Taverner et al., 2000). If educators are going to develop useful new approaches to training health professionals in SA, particularly approaches that emphasize SBIRT principles, they will need educational measurement tools that are consistent with SBIRT treatment approaches, that are standardized across educational settings, and that are appropriate to different points on the educational continuum.

We have previously developed easily administered, survey-based educational measurement tools that can be used by medical educators to assess educational outcomes important to the management of common clinical problems, such as chronic pain and intimate partner violence. This work has been based on standard educational assessment techniques that emphasize measurement of fact-based skills, such as knowledge, attitudes, and self-reported behaviors (KAB). The tools have been shown to be reliable, responsive to educational interventions, correlated with actual office practices, and capable of discriminating experts from non-experts (Short et al., 2006a; Harris Jr. et al., 2008a; Short et al., 2006b; Harris Jr. et al., 2008b). Because they are in the public domain, they can be evaluated and used by others in different settings (Connor et al., 2011).

While such work can facilitate the development and assessment of educational programs that target specific knowledge and attitudes (what the student “knows”), it may not address other important outcomes such as whether the student “knows how,” which Miller defines as “competence” (Miller, 1990). This outcome deals with clinical reasoning, a group of cognitive skills involved in patient evaluation and management (Barrows and Feltovich, 1987). Typical survey-based assessment approaches, particularly multiple-choice questions

and even patient management problems, do not seem to adequately assess clinical reasoning (van der Vleuten and Newble, 1995). More recent work suggests that a different approach, the script concordance test (SCT), is useful for assessing clinical reasoning in a variety of educational settings (Meterissian, 2006; Lambert et al., 2009; Carrière et al., 2009). The SCT uses brief clinical vignettes to present a decision-making problem to the student with a five-category Likert-type answer. The novel feature of the SCT is that scoring is based on a comparison with the answers given by experts, where it is not necessary for all experts to agree on a single response. The rationale of this approach is that the decision-making “scripts” experts use to solve problems may converge, but not be identical; the greater the concordance between a student's answers and the ones chosen by experts, the more expertise the student possesses (Charlin et al., 2000). There is, however, little work to date correlating the SCT with fact-based tests, so it is not yet clear that the SCT measures a unique educational construct (i.e., clinical reasoning expertise vs. knowledge). Additionally, the SCT has not yet been evaluated in the context of SA education.

We sought to develop a reliable and valid survey to measure the SA recognition and management skills of health professionals, within the context of an SBIRT approach. We anticipated that a useful survey, based on nationally-recognized standards, would be internally consistent and would discriminate novice physicians from physicians with additional SA/SBIRT training and expertise.

2. Methods

2.1. Item Development

We convened an expert panel of SA educators and medical educators (listed in Acknowledgements), outlined a goal of a final survey-based educational measurement tool that would contain 50–100 items, explained the KAB and SCT approaches to the panel, and asked the expert group to broadly define the knowledge and skills that a primary care physician should possess to appropriately manage persons with SA disorders using an SBIRT approach. Based on this consensus, two experts (JB, ML) were asked to independently prepare 50–60 draft KAB and 50–60 draft SCT items. The KAB items were either correct or incorrect, but were presented as a 6-point Likert-type scale (1 to 6) reflecting the respondent's degree of confidence in their answer, with the item writer determining the correct answer. We used this scoring approach because it opens up the scoring range and, to the extent that it captures respondent confidence, it may better predict performance (Zelevnik et al., 1988).

We followed the recommendations of Fournier et al. for SCT item development and scoring (Fournier et al., 2008). Basically, SCT items were prepared in three parts, a statement of a problematic clinical situation, followed by a new piece of information, followed by a question, such as whether a diagnostic hypothesis or a clinical approach was more or less appropriate for the particular clinical situation, *considering the new information*. SCT answers were given using a 5-point Likert-type scale (–2 to 2), and were scored using modal scoring approach. This approach allows for divergence in expert answers and credits respondents based on how many experts selected the answer they chose. With this approach, SCT items have a maximum score of 1, assigned to the most common (modal) answer; intermediate scores based on relative frequency to the mode are assigned to other answers selected by the experts; and a score of 0 is assigned to answers not selected by any expert.

To calibrate individual subject scores for comparison purposes, we used a 0–5 range for all items, where “5” was assigned to the extreme “correct” KAB choice, 4 to the next most extreme correct KAB choice, down to 0 for the extreme “incorrect” choice. We scored the SCT items on a 0–5 scale by multiplying the original 0–1 SCT scores by 5†.

2.2. Item Refinement

We used a multistage approach to reduce the number of KAB and SCT items in the final survey. Two independent reviewers (RS, PS) reviewed all items for clarity and relevance. All confusing or irrelevant items were either adjusted or removed, based on consensus of the two reviewers. We next placed the remaining draft survey items on a password-protected Website and asked a group of volunteer physician subscribers to an online continuing medical education (CME) Website (vlh.com) and a group of 17 outside expert physicians recruited by our expert panel (RS, PS, JB) to complete them. Volunteer physician eligibility criteria were that subjects were engaged in the practice of primary care medicine and saw 40+ patients per week. These physicians were asked whether they had undertaken fellowship training or received certification in SA. Volunteer physicians received \$100 in credit from the CME Website. The SA experts were required to be physicians engaged in the active practice and teaching of SA and received a \$200 cash stipend for participation. This activity was reviewed by the Argus Institutional Review Board (Tucson, AZ) and determined to be exempt from Department of Health and Human Services regulations for the protection of human research subjects. After all subjects completed the draft survey, we used data from the 17 physician experts to score the SCT items.

We used a predefined psychometrically-based technical plan to remove items from the draft survey. No items were added. We began by removing items with excessive difficulty for the 17 experts. Specifically, we defined correct KAB answers as either of two extreme Likert scale numbers in the appropriate direction. Thus, for KAB refinement purpose, a “correct” answer to a Likert-type item with six options, where the expert writer determined the correct answer was agreement with the question (versus disagreement), was considered to be a choice of 1= “strongly agree,” or 2 = “agree”). We removed KAB items where <60% of the 17 experts made the correct choice under these conditions (i.e., desired item difficulty index for experts $\geq .6$). We removed SCT items where three or more experts selected both extreme choices (i.e., 3+/17 selected “contraindicated” and 3+/17 also selected “necessary”). An item exclusion criterion based on response congruence has not been generally discussed in the SCT literature, but is reasonable, given that “...optimal SCT questions are those that produce a small range of expert responses clustered around a modal answer” (Lubarsky et al., 2011).

For the next stage of item analysis, we flagged KAB items and SCT items with a point biserial correlation ≤ 0.10 in the group of volunteer physician subjects for removal (Dungan, 1996). The point biserial correlation measures the relationship between the response to an individual item and the overall score on the KAB or SCT subscale. The final decision on item removal was based on expert review (RS, JB).

2.3. Survey Validation

After the draft survey was reduced, we estimated the internal consistency of the final survey and the KAB and SCT subscales using Cronbach's alpha. Cronbach's alpha ranges from 0 to 1, with values greater than 0.7 generally indicating that survey or test items move consistently and appear to be measuring the same hypothetical construct or latent variable. Previous SCT research has used an alpha level > 0.7 as a criterion for high internal consistency (Lubarsky et al., 2011).

Because the data were not normally distributed, we used the Kruskal-Wallis test to compare overall scores and subscale scores among all three groups: expert physicians, non-expert

[†]Further clarification of the scoring method can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

physicians with additional SA training, and non-expert physicians without additional SA training. We used the Mann-Whitney test with Bonferroni correction *post hoc* to further analyze specific between-group differences. In order to estimate whether the KAB and SCT subscales were measuring the same educational construct, we calculated correlation coefficients between the two subscales.

3. Results

The expert consensus determined that KAB and SCT questions should emphasize: (1) screening for SA disorders, (2) interpreting screening results, (3) managing common SA disorders in the outpatient setting, and (4) re-evaluating responses to treatment. With this guidance our expert authors developed 60 draft KAB items and 53 draft SCT items. Following expert review, we removed 20 KAB items and one SCT item, leaving a combined draft survey of 92 items.

The draft survey was offered online between July 23 and September 10, 2010. During that time 100 volunteer physicians and 17 SA expert physicians completed the survey. Survey items were arranged by group, with KAB items preceding SCT items. SCT items were randomized so that items with similar scenarios (stems) were not grouped together. Most study subjects were male allopathic physicians. The most common medical specialty was Internal Medicine (35%-41%). According to self-reported data, 13 volunteer physicians had additional SA certification or fellowship training. The demographic characteristics of the study subjects are shown in Table 1.

Based on the item difficulty criteria, we removed seven of 40 draft KAB items and nine of 52 draft SCT items. Using the point biserial correlation criterion we flagged 4 additional KAB and 10 additional SCT items for potential removal. After expert review, we removed all but one flagged KAB item, "I can use a medical interview to detect problem drinking," and all 10 flagged SCT items. The point biserial correlation for the one retained KAB item was 0.079. The final survey tool, which we have named the Physicians' Competence in SA Test (P-CSAT), contained 63 total items with two 30KAB-item and 33-SCT-item subscales. The steps of item development and refinement are summarized in Figure 1. The final 63 P-CSAT items fell into the following categories: (1) screening for SA disorders, N= 9, (2) interpreting screening results, N= 16, (3) managing common SA disorders in the outpatient setting, N= 29, (4) re-evaluating responses to treatment, N= 9[‡].

The final P-CSAT and its subscales demonstrated good internal consistency. Based on 117 total participants, the Cronbach's alpha was 0.901 for the 63-item P-CSAT, 0.887 for the 30-item KAB subscale, and 0.797 for the 33-item SCT subscale. Based on 100 volunteer physicians, the Cronbach's alpha was 0.839 for the P-CSAT, 0.805 for the KAB subscale, and 0.763 for the SCT subscale.

Kruskal-Wallis tests demonstrated significant differences in overall P-CSAT scores ($\chi^2_{(2)} = 41.930$, $p < 0.001$), KAB subscale scores ($\chi^2_{(2)} = 44.788$, $p < 0.001$), and SCT subscale scores ($\chi^2_{(2)} = 29.696$, $p < 0.001$) among the three physician groups. The mean, median, and standard deviation of P-CSAT scale and subscale scores for each physician group are shown in Table 2. The mean and median scores were in the expected direction, with SA experts scoring the highest and volunteer physicians without any SA training scoring the lowest among the three groups on the P-CSAT and its two subscales.

[‡]A further description of the P-CSAT items and item types can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

Between-group comparisons, using the Mann-Whitney test with Bonferroni correction, demonstrated significant differences in overall P-CSAT scores between volunteer physicians without SA training and SA experts ($Z=6.413$, $p<0.001$, effect size $r=0.629$), and between volunteer physicians with SA training and SA experts ($Z=3.704$, $p<0.001$, effect size $r=0.677$). The difference in P-CSAT scores between volunteer physicians with and without SA training was not significant ($Z=1.430$, $p=0.153$). There were significant differences in KAB subscale scores between volunteer physicians without SA training and SA experts ($Z=6.415$, $p<0.001$, effect size $r=0.629$), between volunteer physicians with SA training and SA experts ($Z=3.433$, $p<0.001$, effect size $r=0.627$), and between volunteer physicians with and without SA training ($Z=2.467$, $p=0.014$, effect size $r=0.247$). For the SCT subscales there were significant differences between volunteer physicians without SA training and SA experts ($Z=5.358$, $p<0.001$, effect size $r=0.525$), and between volunteer physicians with SA training and SA experts ($Z=3.913$, $p<0.001$, $r=0.714$). The difference in SCT scores between the volunteer physicians with and without SA training was not significant ($Z=0.446$, $p=0.656$).

The correlation between the KAB and SCT subscales was moderate and significant. This was true for 100 volunteer physicians ($r=0.495$, $p<0.001$) and for all 117 participants ($r=0.613$, $p<0.001$).

4. Discussion

The key study findings are that the P-CSAT, when used in a group of physicians, is internally consistent and appears to distinguish between physicians with different levels of SA expertise. This easily administered survey may be a useful tool to help educators assess SA competence and measure improvement following training that is based on SBIRT principles.

There are several novel features of the P-CSAT that deserve comment and further evaluation. In order to assess KAB, we did not rely on a typical multiple-choice question format. Instead, we used a Likert-type approach where the answers were dichotomous, but allowed the respondent to express a level of agreement with the answer. We have shown here and elsewhere (Harris Jr., et al., 2008a) that KAB assessment tools based on this approach provide results that correlate with respondent expertise. This approach has the theoretical benefits of opening up the scoring range, which may increase test discrimination, and reflecting respondent confidence, which may better predict performance than dichotomous scoring. The extent to which these theoretical benefits actually accrue, however, deserves further study.

This study expands the literature on the use of the SCT. It is one of the few studies to compare the SCT with a comparable KAB test in a controlled setting. The expected result, assuming that both approaches measure similar but not identical intellectual constructs in the same domain, is that scores from the two subscales would be correlated, but not identical (Lubarsky et al., 2011). Such results were found here, offering support to the claim that SCTs measure a different construct, presumably decision-making, than is measured by fact-based KAB tests. Due to small sample size, factor analysis is untenable in this study. Future studies that involve larger sample size should use confirmatory factor analysis to further evaluate the empirical underlying structure of the measure.

A potential shortcoming of the SCT development process described here is that the SA experts used for SCT calibration were also used for SCT validation, even though excluding an expert from the reference panel in scoring their own SCT measures did not change the results of the study. While this design does not negate the finding that decision-making

skills, as measured by the SCT, were lower in the volunteer physicians, as expected, it is suggested that additional validation studies should be undertaken in different groups of expert and novice physicians, especially since male respondents were overly represented in both our SA expert panel and volunteer physicians.

We have not shown that the P-CSAT or its subscales is responsive to an educational intervention. This is also an important step in the validation of an educational measurement tool. KAB testing is a common approach to evaluating short-term training outcomes, but the SCT has usually been used to assess longer time frames, such as a year of training. We detected significant differences in KAB and SCT scores between physicians with no SA training and SA experts, which are, presumably, due to a long-term training difference where both KAB and decision-making are improved. However, KAB scores between volunteer physicians with no SA training and those with some SA training were significantly different, while SCT scores between these two groups were not. This discrepancy, to some extent, met our expectation that the longer time (or more intense level of training) may be required to change SA decision-making skills compared to knowledge, attitudes, and beliefs. We hypothesize that the P-CSAT will show greater KAB changes than SCT changes following a short-term SA educational intervention, such as continuing medical education, but that changes will be similar following a long-term SA educational intervention.

We developed this assessment tool as part of a larger project to train physicians in SA management. In terms of future evaluation of the P-CSAT, we plan to report the responsiveness of the P-CSAT and its subscales in different groups of physicians following an educational intervention when this work is completed. In the meantime, however, we seek to place the PCSAT in the public domain (see the Supplementary Material) so that others may study it further or use it for formative evaluation purposes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are extremely grateful to Jorge Ruiz, MD, whose guidance on the SCT was pivotal to the development of this research project. We also thank the SA experts who helped develop and refine the P-CSAT: James Berry, DO; Charles Chabal, MD; Robert Hazlett, PhD; Maria Llorente, MD; Richard Saitz, MD, MPH; Paul Seale, MD; and Lawrence Schonfeld, PhD.

Role of Funding Source. The development of the P-CSAT was supported by Small Business Innovation and Research (SBIR) grant R44DA026218 from the National Institute on Drug Abuse. The sponsor had no role in the design and conduct of the study, the collection, management, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript. The Principal Investigator (Dr. Harris) had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

- Barrows HS, Feltovich PJ. The clinical reasoning process. *Med. Educ.* 1987; 21:86–91. [PubMed: 3574170]
- Bernstein E, Bernstein J, Feldman J, Fernandez W, Hagan M, Mitchell P, Safi C, Woolard R, Mello M, Baird J, Lee C, Bazargan-Hejazi S, Broderick K, Laperrier KA, Kellermann A, Wald MM, Taylor RE, Walton K, Grant-Ervin M, Rollinson D, Edwards D, Chan T, Davis D, Buchanan J, Aseltine R, James A, Schilling E, Abu-Hasaballah K, Baumann BM, Boudreaux ED, Maio RF, Cunningham RM, Murrell T, Doezenia D, Anglin D, Eliassen A, Martin M, Pines J, Buchanan L, Turner J, D'Onofrio G, Degutis LC, Owens P. An evidence based alcohol screening, brief intervention and referral to treatment (SBIRT) curriculum for emergency department (ED) providers improves skills and utilization. *Subst. Abuse.* 2007; 28:79–92.

- Bernstein E, Bernstein J, Feldman J, Fernandez W, Hagan M, Mitchell P, Safi C, Woolard R, Mello M, Baird J, Lee C, Bazargan-Hejazi S, Durham BA, Broderick K, LaPerrier KA, Kellermann A, Wald MM, Taylor RE, Walton K, Grant-Ervin M, Rollinson D, Edwards D, Chan T, Davis D, Marshall JB, Aseltine RH Jr, James A, Schilling EA, Abu-Hasaballah K, Harel O, Yan J, Baumann BM, Boudreaux ED, Maio R, Cunningham R, Murrell T, Bauer M, Doezeema D, Anglin D, Eliassen A, Martin M, Pines J, Buchanan L, Turner J, D'Onofrio G, Degutis LC, Owens P. The impact of screening, brief intervention and referral for treatment in emergency department patients' alcohol use: a 3-, 6- and 12-month follow-up. *Alcohol Alcohol*. 2010; 45:514–9. [PubMed: 20876217]
- Carrière B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. *Ann. Emerg. Med*. 2009; 53:647–52. [PubMed: 18722694]
- Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. *Teach. Learn. Med*. 2000; 12:189–95. [PubMed: 11273368]
- Connor PD, Nouer SS, Mackey SN, Tipton NG, Lloyd AK. Psychometric Properties of an intimate partner violence tool for health care students. *J. Interpers. Violence*. 2011; 26:1012–35. [PubMed: 20587479]
- Dungan, L. Examination development. In: Browning, AH.; Bugbee, AC.; Mullins, MA., editors. *Certification: A NOCA Handbook*. National Organization for Competency Assurance; Washington, DC: 1996. p. 1-40.
- Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med. Inform. Decis. Mak*. 2008; 6:8–18.
- Gryczynski J, Mitchell SG, Peterson TR, Gonzales A, Moseley A, Schwartz RP. The relationship between services delivered and substance use outcomes in New Mexico's Screening, Brief Intervention, Referral and Treatment (SBIRT) initiative. *Drug Alcohol Depend*. 2011 epub ahead of print: PMID: 21482039.
- Harris JM Jr, Fulginiti JV, Gordon PR, Elliott TE, Davis BE, Chabal C, Kutob RM. KnowPain-50: a tool for assessing physician pain management education. *Pain Med*. 2008a; 9:542–54. [PubMed: 18266812]
- Harris JM Jr, Elliott TE, Davis BE, Chabal C, Fulginiti JV, Fine PG. Educating generalist physicians about chronic pain: live experts and online education can provide durable benefits. *Pain Med*. 2008b; 9:555–63. [PubMed: 18266811]
- Lambert C, Gagnon R, Nguyen D, Charlin B. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiat. Oncol*. 2009; 4:7. [PubMed: 19203358]
- Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CP. Script concordance testing: a review of published validity evidence. *Med. Educ*. 2011; 45:329–38. [PubMed: 21401680]
- Meterissian SH. A novel method of assessing clinical reasoning in surgical residents. *Surg. Innov*. 2006; 13:115–119. [PubMed: 17012152]
- Miller GE. The assessment of clinical skills/competence/performance. *Acad. Med*. 1990; 65(Suppl.):S63–67. [PubMed: 2400509]
- Office of National Drug Control Policy. *National Drug Control Strategy – 2011*. Washington, DC: 2011. p. 27 available at: <http://www.whitehouse.gov/sites/default/files/ondcp/ndcs2011.pdf>.
- Parish SJ, Ramaswamy M, Stein MR, Kachur EK, Arnsten JH. Teaching about substance abuse with Objective Structured Clinical Exams. *J. Gen. Intern. Med*. 2006; 21:453–459. [PubMed: 16704387]
- Short LM, Alpert E, Harris JM Jr, Surprenant ZJ. A tool for measuring physician readiness to manage intimate partner violence. *Am. J. Prev. Med*. 2006a; 30:173–180. [PubMed: 16459217]
- Short LM, Surprenant ZJ, Harris JM Jr. A community-based trial of an online intimate partner violence CME program. *Am. J. Prev. Med*. 2006b; 30:181–5. [PubMed: 16459218]
- Taverner D, Dodding CJ, White JM. Comparison of methods for teaching clinical skills in assessing and managing drug-seeking patients. *Med. Educ*. 2000; 34:285–91. [PubMed: 10733725]
- van der Vleuten CP, Newble DI. How can we test clinical reasoning? *Lancet*. 1995; 345:1032–1034. [PubMed: 7605439]

Zelevnik C, Hojat M, Goepf CE, Amadio P, Kowlessar OD, Borenstein B. Students' certainty during course test-taking and performance on clerkships and board exams. *J. Med. Educ.* 1988; 63:881–91. [PubMed: 3199416]

Figure 1.
P-CSAT Item Development and Refinement Flow Chart

Table 1

Study Participant Characteristics

Characteristic	Volunteer Physicians	SA Experts
Total	100	17
Males (%)	73 (73.0)	16 (94.1)
Mean Age (SD)	51.5 (9.2)	53.8 (10.7)
MD Degree (%)	92 (92.0)	16 (94.1)
DO Degree (%)	8 (8.0)	1 (5.9)
Primary Board Certification (%)		
Internal Medicine	35 (35.0)	7 (41.1)
Family Medicine	16 (16.0)	6 (35.3)
Emergency Medicine	7 (7.0)	0
Obstetrics & Gynecology	4 (4.0)	0
Pediatrics	4 (4.0)	0
Psychiatry/Neurology	2 (2.0)	4 (23.5)
Other	14 (14.0)	0
None	18 (18.0)	0
Additional Training in SA (%) *	13 (13.0)	17 (100.0)

* Have you had any fellowship training or certification in drug or substance abuse?

Table 2

Median and Mean P-CSAT and KAB and SCT Subscale Scores for Study Participants Based on Level of SA Expertise

Scale (Max Possible Score)	SA Experts Median/Mean Score (SD) (N=17)	Volunteer Physicians with SA Training Median/Mean Score (SD) (N=13)	Volunteer Physicians without SA Training Median/Mean Score (SD) (N=87)
P-CSAT (315)	265.31/260.67 (12.22)	208.16/212.58 (33.55)	202.19/196.22 (29.72)
KAB (150)	129.00/127.24 (8.61)	100.00/106.23 (18.08)	94.00/93.32 (11.16)
SCT (165)	136.27/133.43 (10.11)	110.16/106.35 (19.91)	106.91/102.90 (22.41)

All three-way row comparisons (i.e., overall scale and both subscale score differences between physician groups) are significant ($P<0.001$) via Kruskal-Wallis test. See text for further discussion of between-group differences.