



Published in final edited form as:

J Subst Abuse Treat. 2012 March ; 42(2): 201–212. doi:10.1016/j.jsat.2011.10.017.

Impact of a training and certification program on the quality of interviewer-collected self-report assessment data

Janet C. Titus, (Ph.D.)^{a,*}, Douglas C. Smith, (Ph.D.)^b, Michael L. Dennis, (Ph.D.)^a, Melissa Ives, (M.S.W.)^a, Laine Twanow, (B.A.)^a, and Michelle K. White, (Ph.D.)^c

^aChestnut Health Systems, Normal, IL 61761, USA

^bUniversity of Illinois, Urbana, IL 61801, USA

^cQualityMetric, Inc., Lincoln, RI 02865, USA

Abstract

The purpose of this study was to evaluate the impact of an assessment training and certification program on the quality of data collected from clients entering substance abuse treatment. Data were obtained from 15,858 adult and adolescent clients entering 122 treatment sites across the United States using the Global Appraisal of Individual Needs-Initial (GAIN-I). GAIN Administration and Fidelity Index (GAFI) scores were predicted from interviewer certification status, interviewer experience, and their interactions. We controlled for client characteristics expected to lengthen or otherwise complicate interviews. Initial bivariate analyses revealed effects for certification status and experience. A significant interaction between certification and experience indicates interviewers attaining certification and having more experience far outperformed certified interviewers with low experience. Although some client characteristics negatively impacted fidelity, interviewer certification and experience remained salient predictors of fidelity in the multivariate model. The results are discussed with regard to the importance of ongoing monitoring of interviewer skill.

Keywords

Training; Supervision; Data quality; Assessment; Quality assurance; Monitoring

1. Introduction

Interviewer-administered assessments are routinely used in substance abuse treatment to make important decisions about client diagnosis, placement, and progress. Researchers also depend on data collected with assessments to advance the understanding of addiction and recovery. Because of the magnitude of the decisions that are made using data from interviewer-administered assessments, many treatment and research programs use evidence-based assessments, those with a history of scientific validation and a strong track record of yielding reliable data.

Despite the use of high-quality interviewer-administered assessments, one of the most important parts of the assessment is often overlooked—the interviewer. It is well known from research in epidemiology, sociology, and statistical methods that effects associated

with the interviewer can impact the quality of data and subsequent inferences (Dijkstra & van der Zouwen, 1988; Edwards et al., 1994; Pickery & Loosveldt, 2004). Although interviewers are typically trained in assessment administration, characteristics of training vary widely, and attention to the resulting and ongoing level of skill is often not monitored (Viterna & Maynard, 2002). Data collection procedures used in substance abuse treatment studies are rarely evaluated for their efficacy in ensuring reliable and valid data. This is in sharp contrast to the quality and quantity of ongoing supervision and fidelity monitoring of the clinical delivery of evidence-based treatments reported in the substance abuse treatment field. A recent quality of evidence review on outpatient adolescent treatment programs did not include implementation of data collection as a criterion along which to judge the quality of the evidence (Becker & Curry, 2008).

1.1. Training, supervision, and their relationship to data quality

Available evidence supports training and supervision as effective ways to reduce interviewer-related error, although neither appears to be sufficient by itself. Fowler and Mangione (1983, 1990) randomly assigned trainees to four training programs that varied in time and intensity in combination with three levels of supervision. Analyses focusing on the impact of training alone on interviewer skill (Fowler, 1991) revealed that as length of training increased, percent of errors decreased. The largest reduction in error occurred between the half-day and 2-day trainings; beyond 2 days of training, percentage of errors largely leveled out. Billiet and Loosveldt (1988) observed a similar finding in their study comparing interviewer performance on those exposed to 3 versus 15 hours of training. Interviewers with more training and practice produced more complete data.

Training alone does not appear to be sufficient when it comes to reducing interviewer error and bias. Fowler and Mangione (1983, 1990) found no main effects in a crossed analysis of training by type of supervision, but a significant interaction was detected. The most intensive level of supervision (being taped) produced lower error for those who received the least and the most amount of training. Those who received the most training but *were not* taped had the highest levels of error.

Fowler and Mangione (1990) recommend at least 2- to 3-day trainings that include supervised practice to obtain minimally adequate skills. Furthermore, interviewers should be supervised and regularly monitored using a systematic approach, one in which someone specially trained to monitor listens to actual interviews and systematically rates them along several criteria for conducting quality interviews (Fowler, 2009; Fowler & Mangione, 1990). Rating sheets and systems can be used (e.g., Cannell & Oksenberg, 1988; Mathiowetz & Cannell, 1980), and feedback should be immediate. It is further recommended that all interviews be taped or digitally recorded (e.g., Biemer, Herget, Morton, & Willis, 2000), and the interview to be rated should be randomly chosen. Ongoing direct supervision is needed through taping or observation to make sure good interviewing skills are used. The bottom line: Interviewers who are well trained and monitored perform better, and their better performance translates to better quality data.

1.2. Interviewer experience and data quality

Following training, as interviewers gain experience and become more practiced at the role of “interviewer,” we would expect the quality of their performance to increase. Despite mixed findings, much of the evidence on the impact of interviewer experience on data quality supports the finding that interviewers’ performance becomes less error prone and more efficient with practice.

Olson and Peytchev (2007) studied the effect of interview order (i.e., an interviewer's first vs. 20th interview) and interview experience (1 year or less vs. more than 1 year) on the duration of the interview. They found that within a single project, interviewers picked up speed as they conducted more interviews. In addition, interviewers with more prior experience (more than 1 year) completed interviews significantly faster than new interviewers. For every three interviews conducted by a new interviewer, the total time to administer dropped by an average of three minutes.

In an analysis of measurement error in baseline assessments collected over a 3-year period (Chávez, O'Nuska, & Tonigan, 2009), interviewers became less error prone over time. More errors (i.e., inconsistent answers and missing values) were found in Year 1 than Years 2 and 3, with significantly more missing values in the earlier life of the data collection period. On the other hand, Carton and Loosveldt (1998) found that experienced interviewers obtained more missing data and more "don't know" responses than new interviewers. However, the experienced interviewers also obtained more information from open-ended questions.

1.3. Purpose of study

It is clear that training, supervision, and ongoing practice contribute to interviewer skill, which in turn affects the quality of data. The purpose of this study was to evaluate the impact of an assessment training and certification program on the quality of the data collected from clients entering substance abuse treatment. The assessment is from the Global Appraisal of Individual Needs (GAIN; Dennis, Titus, White, Unsicker, & Hodgkins, 2003) family of semistructured biopsychosocial instruments. We know of no substance abuse treatment studies to date that have systematically examined the quality of data as it relates to interviewer skill obtained through assessment training and certification.

Based on prior research in other areas, we expect to find the following relationships: (a) As interviewers attain assessment certification, the quality of the data will increase; (b) as interviewers become more experienced at conducting the assessment, the quality of the data will increase; (c) the highest quality data will be associated with interaction effects of certification category by experience, such that those certification categories requiring greater administration preparation and ongoing supervision will produce better quality data.

2. Materials and methods

2.1. GAIN-Initial assessment

Data for this study are from the GAIN-Initial (GAIN-I; Dennis, Scott, & Funk, 2003), the intake form of the GAIN assessments. The GAIN-I covers eight life domains, including, but not limited to, substance use, mental health, and legal and vocational areas. Data collected via the 90-minute GAIN-I provide diagnostic impressions based on the American Psychiatric Association's (APA, 1994, 2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (*DSM-IV* and *DSM-IV-TR*) and the American Society of Addiction Medicine (1996, 2001) patient placement criteria (PPC-2, 1996; PPC-2R, 2001). Administration can be conducted via paper and pencil with subsequent data entry or by computer-aided administration. In both adults and adolescents, the main scales have good internal consistency and test-retest reliability and are highly correlated with measures of use from timeline follow-back measures, urine tests, collateral reports, treatment records, and blind psychiatric diagnosis (Dennis et al., 2003; Dennis et al., 2002, 2003, 2004; Godley, Godley, Dennis, Funk, & Passetti, 2002; Shane, Jasuikatis, & Green, 2003). Review copies of the GAIN instruments, manuals, syntax for creating scales and problem-specific group variables, a comprehensive list of supporting studies, detailed psychometrics, and norms for adults and adolescents are publically available at www.chestnut.org/li/gain/.

2.2. GAIN-I training and certification

To become certified as a GAIN-I interviewer, one must complete a rigorous training and quality assurance process. Interviewers are trained to administer the GAIN-I in either a national training or one conducted by a certified local GAIN trainer at their agency.

2.2.1. GAIN-I administration trainings—National GAIN trainings are held at Chestnut Health Systems' GAIN Coordinating Center or at regional locations. National trainings are designed on the “train the trainer” model. That is, several staff from each participating agency or research site attend a national training with the goal of becoming certified to both administer and train their own agency's staff on the assessment. National GAIN trainings are 3.5-day intensive events featuring a comprehensive curriculum, a variety of teaching methods, and a great deal of hands-on practice and feedback. The training is highly interactive, with instruction on the basics of GAIN-I administration demonstrated using sample cases. Trainings typically have 30 to 35 trainees total divided into small groups of no more than 6 individuals to allow for more concentrated instruction and practice administering the GAIN-I.

Local GAIN trainings are conducted by agency staff who have successfully completed a certification program to lead GAIN trainings in their agencies. The local trainings use the same materials and resources and follow the same general training model as those used in the national trainings. Local trainers typically train only a few staff at a time. The trainings typically last 1 to 2 days and involve both didactic instruction with a heavy emphasis on small group practice. More information on GAIN trainings can be found at <http://www.chestnut.org/li/gain/>.

2.2.2. GAIN-I administration training philosophy and curriculum—The GAIN developers believe that assessments relying purely on standardized administration methods jeopardize the validity of answers when respondents do not understand the items or directions (e.g., see Suchman & Jordan, 1990). For all GAIN assessments, standardization of understanding is more important than standardization of administration. The goal is for respondents to *understand* the items as intended so they can offer the most valid response. When the use of standard administration guidelines prevent the respondent from understanding an item as intended or has the potential to derail a smooth interaction, deviations from the standard are expected. Knowing when and how to appropriately deviate from the assessment guidelines is the challenge and the focus of a good portion of GAIN administration training.

The GAIN-I is introduced at training as a semistructured assessment, a cross between a highly structured standardized assessment and a conversational clinical interview. Standardized administration guidelines are introduced at the beginning of training; trainees are then taught to balance standardized administration with aspects of conversational discourse that keep the conversation moving smoothly and logically and communicate that the interviewer is listening to the respondent.

GAIN-I administration topics covered include conducting semistructured interviews, nondirective probing, answering respondent questions, managing the flow of the interview, orienting and teaching the respondent their role, rating treatment urgency and denial/misrepresentation, documenting changes and nonresponse, navigating through complex gridded items, and recording administrative information. Trainees also receive instruction on the quality assurance model and practice applying its guidelines by rating a taped administration with group discussion. Clinical interpretation for diagnosis and placement is covered in a general overview. Before they leave national trainings, trainees can record a

practice interview and receive quality assurance feedback to start their journey to certification.

2.2.3. GAIN-I administration quality assurance model—GAIN-I administration quality assurance (A-QA) is an iterative process that consists of monitoring an interviewer's skills at administering the GAIN-I and providing evaluative feedback. Trainees who submit their audio-recorded client interviews to the A-QA team receive detailed feedback on their level of mastery of the administration. The feedback form is structured around the four topical areas of the A-QA model: documentation (the recording of responses, ratings, and administrative information), instructions (the provision of explanations, directions, and transitional statements to the respondent), items (the delivery and clarification of the items on the assessment), and engagement (the quality of the interaction between the interviewer and the respondent). Quality assurance rating criteria are summarized briefly here.

The *Documentation* A-QA rating criteria address whether the information recorded on the assessment is complete and accurate. Completeness is checked with a visual audit of the record while simultaneously listening to the tape-recorded session.

Criteria under the *Instructions* area monitor the clarity and completeness of directions or orienting information given to the respondent. For example, when introducing the GAIN-I, interviewers must explain the purpose of the assessment, the types of questions that are on the assessment and how they are expected to answer, how long it will take, the need for breaks, and confidentiality. Prior to starting a series of items, interviewers should introduce and define the response choices using the appropriate response card if there is one. If respondents do not understand the instructions or how they are being asked to answer, repeating the instructions and/or explaining concepts in the interviewer's own words is permissible as long as information is accurate and nonleading.

Criteria under the *Items* area contain many of the usual guidelines for conducting standardized interviews. For example, items should be read as written and delivered in the order as they appear on the assessment. Skip-outs should be followed. When respondents give unclear or inconsistent answers or answers that fall outside the defined set of response choices, interviewers should clarify response expectations or probe in a nonleading way. In keeping with the GAIN developers' desire for standardization of understanding, interviewers are expected to answer respondent questions on the meaning of an item, phrase, or word if they do not understand it. Definitions or examples should be clear, not replace the item, and not lead the respondent to answer in a particular way. Interviewers are also permitted to initiate clarification in response to nonverbal behaviors that can signal difficulty, such as a participant taking a long time to answer or facial expressions conveying uncertainty.

Engagement criteria capture the quality of interaction between the interviewer and respondent. Keeping the nature of the interaction within professional yet warm and supportive bounds helps the respondent connect with the process. Interviewers are trained to be sensitive to the respondent's needs, check in with the respondent from time to time, and offer breaks as needed. Encouraging, non-evaluative statements on the interview process can be motivating, especially when an interview lasts a long time. An interviewer's performance in each of the four feedback domains is rated along a 4-point scale: excellent, sufficient, minor problems, and problems. To attain GAIN-I administration certification, the interviewer must attain a rating of "sufficient" or above in each of the four areas in a single, complete interview. This typically happens within four taped submissions and must be accomplished within 3 months posttraining. More detailed information on the A-QA model, criteria, and rating scale is available in the GAIN-I manual at <http://www.chestnut.org/li/gain/>.

2.2.4. GAIN-I certification categories—There are several categories of certification available that are pertinent to the current analyses: GAIN Administration Certification, Local Trainer Certification, and Site Interviewer Certification. To be eligible for *GAIN Administration Certification*, a trainee must attend a national training (train the trainer) event and submit recorded interviews until the above-described mastery levels are achieved. The A-QA team provides detailed written feedback and delivers coaching calls. Trainees must complete GAIN Administration Certification within 3 months of attending GAIN training. Postcertification monitoring is typically provided by local trainers as described below.

Trainees who attain GAIN Administration Certification status are eligible to pursue *Local Trainer Certification*. To attain this category of certification, the local trainer candidate first provides GAIN-I administration training to individuals at their own agency. The trainees each submit an audio-recording of a GAIN-I interview to the local trainer candidate, who prepares written feedback following the A-QA model. The local trainer candidate then submits the written quality assurance feedback and the audio-recorded GAIN interview to the A-QA team at Chestnut Health Systems. The A-QA team conducts a blind quality assurance review and produces an evaluative report comparing Chestnut Health Systems' feedback to that of the local trainer candidate. This report cites the local trainer candidate's skills in providing GAIN administration coaching. A local trainer candidate must demonstrate mastery of A-QA and coaching skills in two stages: first, for an interviewer who requires further GAIN-I administration training, and second, for an interviewer who demonstrates mastery of GAIN-I administration. The local trainer candidate must complete both stages of the local trainer certification process within 6 months posttraining.

The GAIN *Site Interviewer Certification* process supports individual agencies to sustain their use of the GAIN through growth and turnover without having to send additional staff to a national training event. A certified GAIN local trainer provides training and quality assurance feedback to trainee interviewers at their own agency. When the local trainer believes the interviewer's skills merit certification, the local trainer forwards written feedback documenting interviewer mastery to Chestnut Health Systems, where the A-QA team reviews the feedback for verification and, if warranted, grants GAIN-I Site Interviewer Certification status. Site interviewers receive ongoing supervision from their on-site local trainers.

2.3. Analytic data set

The study data set includes GAIN-I (Version 5) records collected from January 2002 through August 2008. All intake data were collected by treatment or research staff at treatment sites across the United States with studies funded by the Center for Substance Abuse Treatment (CSAT) and Office of Juvenile Justice and Delinquency Prevention (OJJDP). All interviewers collecting GAIN-I data from clients during the in-person interviews were trained in GAIN-I administration, and most were certified in some category. Data were collected either under general consent for treatment or informed consent to participate in research under the associated Institutional Review Boards. Data were entered into the GAIN data collection and reporting system (Assessment Building System), either directly through computer administration or after the fact. Chestnut Health Systems' GAIN Coordinating Center (GCC) has data sharing agreements with each participating agency and acts as a data clearinghouse to clean, deidentify, and pool the GAIN data, making it available for secondary research (including for this study).

All data sharing agreements are consistent with the 1996 Health Insurance Portability and Accountability Act Privacy and Security Rules (<http://www.hhs.gov/ocr/privacy/>) and the 2009 Health Information Technology for Economic and Clinical Health Act (<http://healthit.hhs.gov>).

From this initial set ($n = 18,532$), 2,674 records were excluded because they were from sites that had requested exclusion, modified the GAIN-I, had significant data cleaning issues, had less than an 80% follow-up rate, used version 5.1 of the GAIN-I (which is significantly different from later versions), submitted incomplete or partial assessments, or were from sites that contributed less than 10 records total. The remaining data set contains 15,858 records from 122 substance abuse treatment sites.

Two additional data sources were merged into the data set: (a) the interviewer's GAIN-I certification category at the time each assessment was administered, and (b) the total number of data cleaning edits required for each record. Certification categories include "not certified," "GAIN administrator," "local trainer," and "site interviewer." The certification status of the interviewers who conducted every GAIN-I interview was identified using the date they attained certification at any category in conjunction with the date of each interview they conducted. For example, if a given interviewer attained GAIN administrator certification on October 8, 2004, then any assessments conducted by that interviewer from October 8, 2004, to the date of their next category of certification were conducted at the GAIN administrator level. Information on certification dates appeared on interviewers' certification certificates and was stored in a quality assurance database by interviewer. Note that the same individual interviewer could have completed multiple interviews at multiple levels of certification over time. In total, 779 unique individual interviewers conducted the assessments.

Information on the number of data cleaning edits required for each record was provided by the GCC's data team, which is responsible for running routine data cleaning checks on all GAIN records and communicating with sites to facilitate the resolution of data problems. There is a wide range of possible data problems, including missing key variables, unexpected values, and inconsistencies within the same record. The total number of identified problems for each record was matched to the corresponding GAIN-I record in the analytic data set.

2.4. Fidelity indicators

Seven individual indicators of the interviewers' skillfulness at conducting and managing a GAIN-I interview are defined in Table 1. All are directly or indirectly associated with skills learned during GAIN training. As an interviewer's skills develop with further certification and experience, we would expect to see the indicators' values improve, and, by extension, we would expect the quality of the GAIN-I data collected by the interviewer to improve. These "fidelity indicators" index the degree to which an interviewer's skill-based behaviors and interview management skills accurately reproduce the goals of GAIN training.

The distribution of each indicator was examined in an effort to identify cutoffs for high, moderate, and low fidelity (see Table 1, last column). The raw score ranges defining high fidelity (fewer problems) correspond to the optimal values based on training and A-QA guidelines and appear in the first position in Table 1's last column. From there, moderate fidelity is defined up to the 90th percentile of any given indicator and appears in the second position (between the slash marks), whereas low fidelity (more problems) is defined by the top 10% of each indicator's raw score distribution and appears in the third position.

Regarding *interview duration*, it is expected that as interviewers increase in skill and become more familiar with the assessment, the time to conduct the interview will decrease. Interviews lasting up to 90 minutes are considered highest fidelity, whereas those lasting 150 minutes or more are deemed lowest fidelity. Regarding *number of breaks*, interviewers are expected to remain sensitive to the needs of respondents and offer breaks as needed, although often, respondents do not need breaks or wish to continue without one. Having up

to one break was considered highest fidelity, two was considered moderate fidelity, and three or more was coded lowest fidelity. *Context effects* include distractions such as noise, interruptions, telephone calls, and the presence of other people. Because interviewers are expected to provide a quiet, private environment for the assessment, any such effects indicate lower fidelity, with three or more considered lowest. The higher the *number of GAIN edits* (i.e., instances of inaccurate or unclear data) identified after local processing of data and submission to the GCC, the lower the fidelity, with six or more edits in a record considered the lowest. Because interviewers are expected to make every effort to obtain valid responses, the higher the *number of missing or “don’t know” responses* from a pool of 96 items in the main outcome scales, the lower the fidelity, with two or more considered the lowest. Interviewers are also expected to pay attention and remain aware of the logical consistency of the responses and clarify those that do not jibe. When the *number of logical inconsistencies* reaches five, moderate fidelity is expected, with 10 or more inconsistencies considered the lowest fidelity. Finally, GAIN interviewers are expected to stress the confidentiality of responses and manage the assessment session in a way that promotes cooperation on the part of the respondent. At the end of each of the eight sections of the GAIN-I, interviewers provide a rating on the extent to which they believe the respondent was able to understand the items and reply in a forthcoming manner (on a 0 to 4 scale, with 4 indicating outright misrepresentation). The eight ratings are averaged. An average *denial-misrepresentation (D-M) rating* of 1 is considered moderate fidelity, whereas an average of 2 or more is considered the lowest.

The GAIN Administration and Fidelity Index (GAFI; $M = 74$, $SD = 15$, range = 0-100) is a summary measure based on the seven fidelity indicators above. Using the raw score ranges shown in the last column of Table 1, each fidelity indicator was assigned a score, with 2 indicating a raw score in the high fidelity range, 1 indicating a score in the moderate fidelity range, and 0 indicating a score in the low fidelity range. To create the GAFI, the scores for the seven indicators were summed, divided by the maximum range of the sum (i.e., 14), and multiplied by 100—resulting in scores from 0 to 100. Although each of the indicators can, under certain circumstances, reflect issues beyond the interviewer’s control (e.g., a participant’s mental acuity, interviews conducted in jail), the overall pattern of problems represented by GAFI is expected to be associated with the quality of the data.

2.5. Analysis

Data were analyzed in a two-step regression model to predict variation in GAFI. To identify possible confounds, zero-order correlations between GAFI scores with assorted measures of client characteristics and level of care placement were inspected for significant relationships. Those with significant correlations ($p < .05$) were entered into the regression equation on Step 1 to control for individual client differences at intake. In Step 2, staff certification category (not certified, GAIN administrator, local trainer, site interviewer), GAIN-I interviewing experience level (less than 15 GAIN interviews completed, 15 or more GAIN interviews completed), and their interactions were entered to determine the additional variance accounted for by these variables. The cut point of 15 interviews completed was used to index experience based on results of an earlier analysis, which showed 15 interviews was the point at which interviewers’ learning curve slope began to smooth out (White, 2006), indicating further experience did not increase data quality as dramatically. Regression coefficients were computed at Steps 1 and 2, followed by the regression coefficient indexing the change in percent of GAFI variance accounted for between Steps 1 and 2. Standardized beta coefficients associated with the levels of interaction between certification and experience were computed at Step 2.

Note that the basics of conducting a GAIN assessment are the same regardless of mode of administration (paper-and-pencil vs. computer aided). All trainees learn to interact with the

client, manage the interview, collect data without jeopardizing reliability and validity, and use the software that allows for data entry post-paper-and-pencil administration or computer administration. Mode of administration of the GAIN-I was not a variable in this analysis.

3. Results

3.1. Descriptive statistics and reliability of GAFI

The distribution of the GAFI composite measure shows that on average, GAIN interviewers are adhering to the quality standards taught during the training and certification process ($M = 74$, $SD = 15$). Approximately 33% of interviews were conducted with a GAFI score of 80% or above, 37% were in the 70s, and 30% of the interviews were conducted with a GAFI score below 70%. As a composite measure, GAFI explains 93% of the variance in the joint distribution of the seven indicators and 16 to 32% of the variance for the seven individual indicators. Thus, it is a reliable and efficient summary measure.

3.2. Bivariate associations

Fig. 1 shows average GAFI z -scores (standard deviations from the mean) for groups differing in certification category and experience prior to controlling for potential confounds such as client and environmental characteristics. The omnibus F test reveals significant main effects for certification category, Cohen's $f = .17$; $F(3, 15,421) = 234.9$, $p < .001$, with post hoc tests revealing significant improvement between not certified ($M = 70.1$) and GAIN administrator ($M = 73.9$) and further improvement for local trainer ($M = 77.2$) and site interviewer ($M = 77.0$). Contrasts between site interviewer and local trainer are not statistically significant. On average, GAFI scores are also significantly higher for the experienced versus inexperienced interviewers, Cohen's $f = .33$; $F(1, 15,419) = 1,911.6$, $p < .001$. The interaction of certification category and experience is also significant, Cohen's $f = .28$; $F(3, 15,415) = 295.9$, $p < .001$, with trend plots suggesting that fidelity increases as rigor of certification preparation increases (i.e., moving from not certified to GAIN administrator/site interviewer to local trainer) within both the low- and high-experience groups. Post hoc tests reveal three contrast groups: low experience not certified/GAIN administrator (Group 1, lowest fidelity), low experience local trainer site interviewer (Group 2, moderate fidelity), and high experience regardless of certification category (Group 3, highest fidelity). Average GAFI scores between these three groups are significantly different. In summary, a small to moderate main effect for certification status and a moderate to large main effect for experience were observed, with noncertified and less experienced interviewers (i.e., 15 completed GAIN interviews) having lower GAFI scores on average. The interaction between experience and certification status was also significant.

3.3. Multivariate regression models

Table 2 presents findings from a series of regression analyses in an effort to explain variance in GAFI scores. Client characteristics—including level of care placement—have the potential to impact the conduct of an assessment session and, by extension, GAFI scores. To examine the effects associated with these variables, we individually regressed a series of client characteristics and level of care placement status on GAFI scores. As shown in the top part of Table 2 (under “Bivariate”), all but two of the variables—age and gender—emerged as statistically significant predictors of GAFI. Specifically, membership in a non-White minority group, higher cognitive impairment, increased global mental health severity, current criminal justice system involvement, greater number of prior drug treatment episodes, and increased utilization of lifetime health services were associated with lower GAFI scores. Two variables, being in school or work and current involvement with the child welfare system, showed modest positive associations with GAFI scores. In addition, lower GAFI scores were associated with clients from residential treatment settings. Considering

that GAFI scores range from 0 to 100, the magnitude of these associations appear modest, with the strongest bivariate predictors of GAFI scores being General Individual Severity Scale scores ($b = -0.29, p < .001$), increased number of prior substance use disorder treatment episodes ($b = -0.24, p < .001$), and residential treatment status ($b = -0.15, p < .001$).

The set of significant predictors from the bivariate analysis was entered into a multivariate regression model (Step 1) to examine their conjoint predictive influence on GAFI scores. This was done in an effort to identify variance associated with potentially confounding variables and distinguish it from variance associated with the variables of interest (i.e., certification category and experience). Although this model is not the focus of this article and its coefficients are not shown in Table 2, it accounted for 14% of the variance in GAFI scores, $df(10, 15,383), p < .001$.

In a second multivariate regression model (Step 2), all of the significant individual variables from Step 1 were retained, and the interaction terms between certification category and experience were entered into the model. (Certification category and experience were not entered individually into the model because their variance is already accounted for in the interaction terms.) Prediction weights and significance levels for the variables in the final model (Step 2) appear in the “Multivariate” column of Table 2. All but one of the interaction terms between certification and experience (that for low experience GAIN administrators) were stronger predictors of GAFI scores when compared with the predictive power of client characteristics and level of care. Furthermore, the magnitude of prediction weights for the interaction terms roughly follow the rank order we would expect for level of experience and training intensity (i.e., certification category attained). This final model accounted for 25% of the variance in GAFI scores, $R^2 = .25, df(17, 15,376), p < .001$; R^2 change = .11, $df(7, 15,376), p < .001$.

4. Discussion

4.1. Summary and interpretation

Looking broadly across interviewer certification categories and experience, the GAIN-I interviewers appear to have done a good job at implementing the interviewing skills they learned in training, with approximately two thirds of the interviews performed at or above a GAFI score of 70. Initial bivariate analyses show that both certification and experience matter. Interviewers who were certified at some level of GAIN administration collected data with greater fidelity than those who were not certified. Interviewers with a history of at least 15 completed GAIN-I interviews collected data with greater fidelity than those with less experience. Interaction effects revealed that certification matters more when interviewers have low experience; once they gain experience, differences between the certification groups diminish.

Several client factors were found to be associated with the fidelity of an interview, such that clients with a more troubled profile (e.g., elevated mental health problems, prior substance abuse treatment, and residential placement) were significantly associated with depressed GAFI scores. Thus, a multivariate analysis was performed controlling for potentially confounding effects associated with clients. Compared with uncertified low experience interviewers, all other certification-by-experience groups collected significantly higher fidelity data. Among the low-experience groups, site interviewers and local trainers performed similarly and stronger than the GAIN administrators. Among the high-experience groups, all certification statuses performed better than those with low experience, with site interviewers showing the strongest results. This pattern of results is similar to that observed

in the bivariate analyses. Confounding effects alone explained 14% of the variance in GAFI scores, with the interaction terms explaining an additional 11%.

Clearly, both certification and experience of interviewers are important to the quality of data, although in this study, certification appears to have had a stronger effect among interviewers with low experience. Among interviewers with high experience, even the noncertified interviewers collected greater fidelity data than those certified with low experience, although their performance was the weakest of those with high experience. This observation begs the question—Can experience take the place of certification? We do not believe so. Gaining experience takes time. It is possible that interviewers with continuing hands-on training and supervision through the certification process will collect greater fidelity data sooner than those who depend only on improving through experience. In addition, interviewers with high experience who achieve some level of certification collect better fidelity data than those without certification.

The pattern of results can be explained by a combination of hands-on training and ongoing supervision coupled with a case flow steady enough to provide experience. Certified GAIN administrators reach certification following 3 months of intensive training and quality assurance monitoring. Those who will be responsible for training and supervising individuals at their own site move to a more advanced level on the certification ladder, that of local trainer. This certification status requires an additional 3 months of intensive hands-on training with one-on-one supervision and coaching. It is important to remember that all local trainers were once GAIN administrators. The superior performance of the local trainers compared with that of the GAIN administrators observed in this study reflects the natural progression of the growth in skills as one successfully masters the requirements of local trainer certification.

What should we make of the findings that site interviewers performed similarly to local trainers at the low level of experience and superior to them at the high level of experience? Recall that site interviewers are trained and monitored by the local trainers at their site. Thus, they receive the most support, maintenance training, and mentoring of all certified interviewers through the ongoing supervision of an on-site local trainer for the entirety of their time as GAIN interviewers. As observed in Fowler and Mangione's classic work (1983, 1990), this combination of hands-on training and ongoing supervision produces the best quality data. Local trainers are mentored and supervised only through their certification as local trainers. As they are typically in administrative roles, they may not have as many opportunities to conduct assessments as the site interviewers, who are often frontline staff. Over time, as the site interviewers accrue more and more experience, their skills continue to grow and they produce data that is superior even to their local trainers. Somewhat paradoxically, it may be that the local trainer's interview quality diminishes over time without continued practice or supervision of their own, despite their competence at supervising others' GAIN administrations. The local trainers may be balancing administrative roles while continuing to interview more sporadically, resulting in lower GAFI scores relative to their progeny.

A similar explanation could be offered for site interviewers' superior performance over GAIN administrators. In this sample, 17% of the assessments were conducted by GAIN administrators, many of whom would not have progressed to local trainer status. GAIN administrators may or may not have the luxury of ongoing supervision postcertification, as their preparation is a one-time, hands-on training with intensive supervision only through certification as a GAIN administrator. The benefits of having a case flow steady enough on which to gain experience coupled with ongoing mentoring and supervision appear to have a powerful effect on the quality of data.

4.2. Strengths and limitations

This study has several strengths, including a large sample size, the use of standardized measures, the inclusion of multiple sites, and multiple sources of data to index data quality (i.e., from training records, GAIN interviewer impressions, data cleaning records, and client responses). Several categories of interviewer certification were compared, including an uncertified group that acted as a control.

There are, however, some important limitations that should be acknowledged. First, although this analysis is based on data from staff of various backgrounds from a wide range of agencies across the United States, it cannot be assumed that this group is representative of all people who will implement the GAIN. The analysis did not consider the impact of all possible staff or agency variations that may play a role in maintaining fidelity in GAIN administration; factors such as the amount of training and supervision led by local trainers, agency climate, and administrative support for implementation are likely to vary across agencies. Secondly, local trainer candidates are responsible for providing training to their agency staff. Although they are free to use resources provided by the GCC, they have a great deal of latitude in the training they provide to site interviewer candidates. Given the performance of the site interviewers in the current analyses, the local trainers appear to have done a fine job with training. However, ultimately, the comparability of training for GAIN administrators and site interviewers cannot be assessed.

4.3. Implications for practice and policy

Our findings support employing the train the trainer model to facilitate widespread use of evidence-based assessments in treatment agencies. When coupled with experience, GAFI scores increased more with certification. However, site interviewers, those who were trained to certification standards by local trainers who had attended a national GAIN training, outperformed all other certification groups at high levels of experience. As site interviewers completed nearly 27% of the interviews in this study ($n = 4,227$), this may be viewed with substantial optimism regarding the acceptable data quality and cost savings that may be achieved with train the trainer models. That is, because site interviewers are likely to be in frontline entry level positions, it is encouraging that they can achieve high GAFI scores after receiving training from local trainers and accruing adequate experience. Interviews conducted by local trainers incur higher costs (i.e., higher salaries for administrative, supervisory positions) with no resulting increase in data quality. In sum, the finding that experienced site interviewers had the highest data quality among certification groups supports the feasibility of the train the trainer model.

These findings also provide insight for key components of assessment training. The quality of data collected increased in relationship to the amount of one-on-one, in-person training and feedback interviewers received through the certification process. As the field moves toward greater utilization of e-learning environments and remote training opportunities in efforts to save costs and achieve widespread implementation, trainers should contemplate ways to maintain the benefits of working closely on an ongoing basis with a trainee who happens to be at a distance.

Numerous variables associated with multiproblem clients (e.g., high psychiatric comorbidity, high service utilization) were associated with lower GAFI scores, indicating that some clients may be more difficult to interview than others. However, certification and experience can protect against effects associated with a more challenging client given their interaction was a much stronger predictor of GAFI scores in the full multivariate model. Training should emphasize the personal control that interviewers have to improve data

quality and deal more effectively with multiproblem clients by participating in the certification process.

Administrators and other agency personnel administering assessments may not see the value of increased data quality as measured here by the GAFI. Reduced administration time, one component of GAFI, has the most direct bearing on implementation costs, but what of the other components of GAFI? It is possible that other items in GAFI also incur costs and impact treatment processes, which should be elaborated in future research and explained to practitioners during training. For example, logical inconsistencies were varied in the current measure (e.g., conflict between days abstinent and days sober, administration time reported was longer than the difference between reported stop and start times). It is reasonable to assume that those inconsistencies that lead to fewer clinical referrals or ineligibility for services will be viewed as more concerning to practitioners, whereas others will be viewed as relevant only to researchers. Interviewers are trained to clarify inconsistencies in time frames of substance use disorder criteria. A simple reporting error may result in no past-year diagnosis, which could have been remedied if interviewing procedures were followed to clarify such an inconsistency. In addition, if gross and unresolved inconsistencies in frequency of substance use appear and result in an inaccurate assessment, a client may be placed in an inappropriate level of care. Additional research may focus on which data quality measures are most important to clinicians, which may increase their enthusiasm for participating in rigorous training and supervision procedures like those described here.

Finally, the data in this study may guide the development of performance benchmarks for trainees in various certification tracks and at different experience levels. Such benchmarks can be used by supervisors to identify both exemplary and subpar performance and could be used during coaching sessions. The GCC currently distributes quarterly reports to help program managers track their participant recruitment and retention benchmarks, which contrasts their performance relative to that of other sites in their grantee cohort. A similar procedure could be adopted for interviewer performance and the implied quality of the data.

Acknowledgments

Analysis of the GAIN data reported in this article was supported by Substance Abuse and Mental Health Services Administration's CSAT under Contract 270-07-0191 using data provided by 122 grantees: TI13190, TI13305, TI13308, TI13313, TI13322, TI13323, TI13344, TI13345, TI13354, TI13356, TI13601, TI14090, TI14188, TI14189, TI14196, TI14252, TI14261, TI14267, TI14271, TI14272, TI14283, TI14311, TI14315, TI14376, TI15413, TI15415, TI15421, TI15433, TI15438, TI15446, TI15447, TI15458, TI15461, TI15466, TI15467, TI15469, TI15475, TI15478, TI15479, TI15481, TI15483, TI15485, TI15486, TI15489, TI15511, TI15514, TI15524, TI15527, TI15545, TI15562, TI15577, TI15584, TI15586, TI15670, TI15671, TI15672, TI15674, TI15677, TI15678, TI15682, TI15686, TI16386, TI16400, TI16414, TI16904, TI16928, TI16939, TI16961, TI16984, TI16992, TI17046, TI17070, TI17071, TI17334, TI17433, TI17434, TI17446, TI17475, TI17476, TI17484, TI17486, TI17490, TI17517, TI17523, TI17535, TI17547, TI17589, TI17604, TI17605, TI17638, TI17646, TI17648, TI17673, TI17702, TI17719, TI17724, TI17728, TI17742, TI17744, TI17751, TI17755, TI17761, TI17763, TI17765, TI17769, TI17775, TI17779, TI17786, TI17788, TI17812, TI17817, TI17825, TI17830, TI17831, TI17864, TI18406, TI18587, TI18671, TI18723, TI19313, TI19323, and 655374. Additional support came from NIAAA grant no. 1K23AA017702-01A2 (Smith). The opinions expressed here belong to the authors and are not official positions of the government.

References

- American Psychiatric Association (APA). Diagnostic and statistical manual of mental disorders. 4th ed.. Author; Washington, DC: 1994.
- American Psychiatric Association (APA). American Psychiatric Association diagnostic and statistical manual of mental disorders. 4th ed.. American Psychiatric Association; Washington, DC: 2000. text revision

- American Society of Addiction Medicine. (ASAM). Patient placement criteria for the treatment of psychoactive substance use disorders. 2nd ed.. Author; Chevy Chase, MD: 1996.
- American Society of Addictive Medicine (ASAM). Patient placement criteria for the treatment of psychoactive substance disorders. 2nd ed.. Author; Chevy Chase, MD: 2001. revised
- Becker SJ, Curry JF. Outpatient interventions for adolescent substance abuse: A quality of evidence review. *Journal of Consulting and Clinical Psychology*. 2008; 76:531–543. [PubMed: 18665683]
- Biemer, P.; Herget, D.; Morton, J.; Willis, G. The feasibility of monitoring field interview performance using Computer Audio Recorded Interviewing (CARI); Paper presented at the 55th annual conference of the American Association for Public Opinion Research and World Association for Public Opinion Research; Portland, Oregon. 2000;
- Billiet J, Loosveldt G. Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly*. 1988; 52:190–211.
- Cannell, C.; Oksenberg, L. Observation of behaviour in telephone interviews. In: Groves, RM.; Biemer, PN.; Lyberg, LE.; Massey, JT.; Nichols, WL., II; Waksberg, J., editors. *Telephone survey methodology*. John Wiley; New York: 1988. p. 475-495.
- Carton, A.; Loosveldt, G. Interviewer selection and data quality in survey research; Paper presented at the 53rd annual conference of the American Association for Public Opinion Research; St. Louis, MO. 1998;
- Chávez, R.; O’Nuska, M.; Tonigan, JS. The evaluation of data quality control in a large 5-year clinical trial; Poster presentation at the 32nd annual scientific meeting of the Research Society on Alcoholism; San Diego, CA. 2009;
- Dennis ML, Godley SH, Diamond G, Tims FM, Babor T, Donaldson J, et al. The Cannabis Youth Treatment (CYT) Study: Main findings from two randomized trials. *Journal of Substance Abuse Treatment*. 2004; 27:197–213. [PubMed: 15501373]
- Dennis ML, Scott CK, Funk R. An experimental evaluation of recovery management checkups (RMC) for people with chronic substance use disorders. *Evaluation and Program Planning*. 2003; 26:339–352.
- Dennis ML, Titus JC, Diamond G, Donaldson J, Godley SH, Tims F, et al. The Cannabis Youth Treatment (CYT) experiment: Rationale, study design, and analysis plans. *Addiction*. 2002; 97(Supp 1):16–34. [PubMed: 12460126]
- Dennis, M.L.; Titus, J.C.; White, M.; Unsicker, J.; Hodgkins, D. Global Appraisal of Individual Needs (GAIN): Administration guide for the GAIN and related measures. Version 5 ed.. Chestnut Health Systems; Bloomington, IL: 2003. Accessed on 12/10/2003 from <http://www.chestnut.org/LI/gain/index.html>
- Dijkstra, W.; van der Zouwen, J. Types of inadequate interviewer behavior in survey-interviews. In: Saris, WE.; Gallhofer, IN., editors. *Sociometric research, Volume 1: Data collection and scaling*. St. Martin’s Press; New York: 1988. p. 24-35.
- Edwards S, Slattery ML, Mori M, Berry TD, Caan BT, Palmer P, et al. Objective system for interviewer performance evaluation for use in epidemiologic studies. *American Journal of Epidemiology*. 1994; 140:1020–1028. [PubMed: 7985650]
- Fowler, FJ. *Survey research methods*. 4th ed.. Sage Publications, Inc; Thousand Oaks, CA: 2009.
- Fowler, FJ. Reducing interviewer-related error through interviewer training, supervision, and other means. In: Biemer, PP.; Groves, RM.; Lyberg, LE.; Nathiowetz, NA.; Sudman, S., editors. *Measurement errors in surveys*. Wiley; New York: 1991. p. 259-278.
- Fowler, FJ.; Mangione, TW. Proceedings of the Survey Research Methods Section, American Statistical Association. 1983. The role of interviewer training and supervision in reducing interviewer effects on survey data; p. 124-128.
- Fowler, FJ.; Mangione, TW. *Standardized survey interviewing: Minimizing interviewer-related error*. Sage; Newbury Park, CA: 1990.
- Godley MD, Godley SH, Dennis ML, Funk R, Passetti L. Preliminary outcomes from the assertive continuing care experiment for adolescents discharged from residential treatment. *Journal of Substance Abuse Treatment*. 2002; 23:21–32. [PubMed: 12127465]

- Mathiowetz N, Cannell C. Coding interviewer behavior as a method of evaluating performance. *Proceedings of the Survey Research Methods Section, American Statistical Association.* 1980:525–528.
- Olson K, Peytchev A. Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly.* 2007; 71:273–286.
- Pickery J, Loosveldt G. A simultaneous analysis of interviewer effects on various data quality indicators with identification of exceptional interviewers. *Journal of Official Statistics.* 2004; 20:77–89.
- Shane P, Jaisiukaitis P, Green RS. Treatment outcomes among adolescents with substance abuse problems: The relationship between comorbidities and post-treatment substance involvement. *Evaluation and Program Planning.* 2003; 26:393–402.
- Suchman L, Jordan B. Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association.* 1990; 85:232–241.
- Viterna, J.; Maynard, DW. How uniform is standardization? Variation within and across survey research centers regarding protocols for interviewing. In: Maynard, DW.; Houtkoop-Steenstra, H.; Schaeffer, NC.; van der Zouwen, J., editors. *Standardization and tacit knowledge: Interaction and practice in the survey interview.* John Wiley & Sons; New York: 2002. p. 365-397.
- White, M. The effects of a quality assurance (QA) and certification program on the quality and validity of adolescent substance abuse treatment data; Presentation at the American Evaluation Association annual conference; Portland, Oregon. 2006;

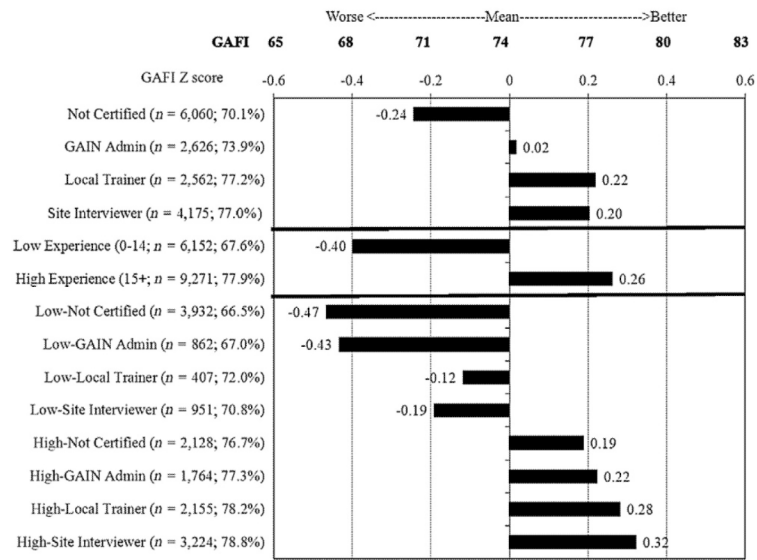


Fig. 1. Average GAFI z-scores by interviewer certification category, experience, and their interactions.

Table 1

Definitions of fidelity indicators

Fidelity indicators	Data source ^a	Raw score ranges ^b
Interview duration—The duration of the interview (in minutes).	GAIN-I	37–90/91–149/150+
No. of Breaks—Number of breaks taken during the administration of the GAIN-I.	GAIN-I	0–1/2/3+
Context effects—13-item (yes/no) index where higher scores indicate increasingly more difficulty with providing or maintaining a quiet and private environment for the interview, respondent attitude or behavioral problems, and ability to complete the interview.	GAIN-I	0/1–2/3+
No. of GAIN edits—Number of edits identified by the GCC data management staff for each GAIN administration. Values are generated via data cleaning syntax used to identify missing values for required data items, common inconsistencies, out of range responses, and other anomalies for each record.	DC	0–1/2–5/6+
No. of missing/don't know—Number of “missing” or “don't know” responses (out of 96 items indexing main outcome variables).	GAIN-I	0/1/2+
No. of logical inconsistencies—Total number of logical inconsistencies (out of 79 identified) present for each GAIN-I record. For example, a recency item (e.g., last time used any substance) should not have a lower valued response (i.e., be longer ago) than the most recent individual recency item (e.g., last time used a specific substance).	GAIN-I ^c	0–4/5–9/10+
D-M ratings—Ratings based on the interviewer's perception (0 = none, 4 = misrepresentation) of the respondent's ability to both understand interview questions and reply in a forthcoming manner. This measure is the average of the D-M ratings for each of the eight sections of the GAIN-I, with higher scores indicating increasing perception that the respondent's answers are not completely truthful.	GAIN-I	0/1/2–4
GAFI—Total fidelity score (range = 0–100) calculated as the sum of the 7 recorded raw fidelity indicator scores (i.e., recoded as 2 = high fidelity/fewer problems, 1 = moderate fidelity, and 0 = low fidelity/more problems using the raw score ranges in the third column of this table) divided by 14 (maximum range of sum) and multiplied by 100.	Calculated from recoded raw fidelity indicator scores	–

^aGAIN-I = interviewer's self-report during GAIN-I administration; DC = Values are from records maintained by the GCC during the data cleaning process.

^bRecoded scores for the raw score ranges are as follows: 2 = high fidelity (fewer problems), 1 = moderate fidelity, 0 = low fidelity (more problems).

^cNo. of inconsistencies is calculated using code based on input from the A-QA team on pairs of item with responses that are not consistent.

Table 2

Bivariate and multivariate predictors of GAFI scores ($N = 15,423$)

Client characteristics	Distribution		Bivariate		Multivariate	
	<i>M</i>	<i>SD</i>	<i>B</i>	<i>p</i>	<i>b</i>	<i>p</i>
Intercept	—	—	—	—	76.3	.00
Female (0–1)	0.26	0.44	-0.01	.23	—	—
Age (10–67)	16.26	3.35	-0.01	.31	—	—
Non-White (0–1)	0.60	0.50	-0.07	.00	-1.79	.00
Cognitive impairment (0–22) ^a	3.42	3.14	-0.07	.00	-0.29	.00
General Individual Severity Scale (0–113) ^b	32.56	21.97	-0.29	.00	-0.15	.00
Current justice system involvement (0–1) ^c	0.73	0.45	-0.09	.00	-2.19	.00
Current child welfare system involvement (0–1) ^d	0.17	0.33	0.04	.00	-1.45	.00
Currently in school or work (0–1)	0.94	0.23	0.05	.00	1.10	.02
No. of prior AOD treatment episodes (0–5+)	0.57	1.06	-0.24	.00	-0.02	.00
Lifetime service utilization (0–1,305) ^e	8.59	19.63	-0.10	.00	-2.30	.00
Level of care placement	<i>n</i>	%	<i>B</i>	<i>p</i>	<i>b</i>	<i>p</i>
Outpatient (0–1)	12538	79.1	—	—	—	—
Residential (0–1)	1817	11.5	-0.15	.00	-3.50	.00
Outpatient continuing care (0–1)	1503	9.5	0.00	.59	0.70	.06
Staff characteristics	<i>n</i>	%	<i>B</i>	<i>p</i>	<i>b</i>	<i>p</i>
Certification category						
Not certified (0–1)	6368	40.2	—	—	—	—
GAIN administrator (0–1)	2689	17.0	0.00	.33	—	—
Local trainer (0–1)	2574	16.2	0.13	.00	—	—
Site interviewer (0–1)	4227	26.7	0.10	.00	—	—
Experience						
<15 interviews (0)	6021	40.7	—	—	—	—
15+ interviews (1)	9402	59.3	-0.33	.00	—	—
Certification by experience						

Client characteristics	Distribution		Bivariate		Multivariate	
	M	SD	B	p	b	p
Not certified (0-1)	4192	26.4	-	-	-	-
GAIN admin (0-1)	883	5.6	-0.11	.00	1.25	.01
Local trainer (0-1)	409	2.6	-0.05	.00	5.33	.00
Site interviewer (0-1)	972	6.1	-0.02	.01	4.68	.00
Not certified (0-1)	2176	13.7	0.08	.00	8.76	.00
GAIN admin (0-1)	1806	11.4	0.08	.00	10.35	.00
Local trainer (0-1)	2165	13.7	0.17	.00	11.56	.00
Site interviewer (0-1)	3255	20.5	0.12	.00	12.95	.00
Model R ² (Step 1) ^f					0.14	.00
Model R ² (Step 2) ^g					0.25	.00
R ² change					0.11	.00

Note. Bold values = $p < .05$.

^aCognitive impairment is based on the Short-Blessed Scale (Items A2a to A2f on the GAIN-1).

^bGeneral Individual Severity Scale is a composite scale on the GAIN-I computed as a total past-year symptom count across 15 scales: Substance Dependence Scale, Substance Abuse Index, Substance Issues Index, Somatic Symptom Index, Depressive Symptom Scale, Homicidal-Suicidal Thought Scale, Anxiety Symptom Scale, Traumatic Symptom Scale, Inattentive Disorder Scale, Hyperactivity-Impulsivity Scale, Conduct Disorder Scale, General Conflict Tactics Scale, Property Crime Scale, Interpersonal Crime Scale, and Drug Crime Scale.

^cCurrent criminal justice involvement is based on indicators including awaiting charges, a trial, or sentencing; in jail/prison or detention; on probation, parole, treatment/work/school release; out on bail; assigned to a sentencing alternative; under house arrest or other forms of court supervision; any other involvement in the criminal justice system.

^dCurrent child welfare system involvement is based on current custody, current living situation, and referral source.

^eLifetime service utilization is the sum of lifetime treatment episodes for substance abuse, mental health, or physical health (including the use of medication).

^fIn this model, GAFI score is regressed on client characteristics and level of care placement variables that were significantly associated with GAFI.

^gIn this model, GAFI score is regressed on the client characteristics and level of care placement variables from step one along with the interaction terms of certification category by interviewer experience.