Journal of Counseling Psychology
2016, Vol. 63, No. 3, 278–293

0022-0167/16/$12.00 http://dx.doi.org/10.1037/cou0000147

Enhancing Feedback for Clinical Use: Creating and Evaluating Profiles of Clients Seeking Counseling

Samuel S. Nordberg, Louis G. Castonguay, Andrew A. McAleavey, Benjamin D. Locke, and Jeffrey A. Hayes
Pennsylvania State University

The current study explored the reliability and clinical utility of a method designed to identify latent classes of students seeking counseling, based on 8 symptom domains and their interactions. Participants were over 50,000 college students in counseling, assessed with the CCAPS-62 and -34 as part of routine clinical care. Latent profile analysis was used to group an exploratory and confirmatory sample of students by reported symptoms across the 8 CCAPS subscales. Profiles were evaluated for reliability and clinical utility, in particular for risk assessment and the prediction of treatment duration and success. Nine reliably stable latent profiles, or groups of profiles, emerged from analysis. Profiles differed significantly in reported symptoms, demographic makeup, psychosocial history, and diagnoses. Additionally, profiles appeared to capture meaningful differences between clients that had implications for relative risk of suicide, self-harm, and violence toward others as well as significant differences in the number of sessions in treatment and the effect size of treatment. Latent profiles of patients appear to capture meaningful, stable differences that could be implemented in an automated system of evaluation and feedback, and that might be useful to clinicians, administrators, and researchers.

Keywords: latent profile analysis, assessment, psychotherapy outcome, informatics, college counseling

Supplemental materials: http://dx.doi.org/10.1037/cou0000147.supp

Measuring symptoms before and during therapy can provide useful and easily interpreted information to enhance clinical judgment about a particular case (Boswell, Kraus, Miller, & Lambert, 2015). There is indeed growing evidence that outcome monitoring in naturalistic settings allows therapists to infer characteristics of a new client, using existing data from similar, previously treated, clients (e.g., Lutz et al., 2005). There is additional evidence that brief self-report instruments can offer some prediction of treatment course (e.g., Lutz, Martinovich, Howard, & Leon, 2002; Stulz, Lutz, Leach, Lucock, & Barkham, 2007). One potential limitation of these types of prediction, however, is that they are often based on a single scale, or a total score. Individuals seeking psychotherapy can be markedly different from one another, and unless those differences are meaningfully captured, there is a risk of overgeneralizing conclusions about patients. For example, describing patients based on a measure of substance use may miss important differences between patients. Are they elevated on substance use because they engage in significant social drinking, or are they coping with or masking other problems, such as depression or anxiety? How concerned should a clinician be by a patient reporting alcohol consumption in the 95th percentile? Often, what is missing is the context in which a particular set of symptoms occur; a single measure does not adequately describe a person.

Indeed, studies have indicated that patients who appear to have similar distress at the start of treatment can have markedly different recovery trajectories, which may result from the influence of other unmeasured variables (Lutz, Stulz & Kock, 2009; Nordberg, Castonguay, Fisher, Boswell & Kraus, 2014; Stulz et al., 2007). In addition, recent advances in the dose–response literature have indicated that patients can require significantly different treatment durations before recovering and ending their treatment (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009; Stulz, Lutz, Kopta, Minami, & Saunders, 2013; Owen, Adelson, Budge, Kopta, & Reese, 2014). It follows that a data-driven method that neglects interactions between measures of symptoms and functioning may fail to capture important differences in intake characteristics that influence expected treatment outcome, and be less sensitive to changes in particular domains (McAleavey, Nordberg, Kraus, & Castonguay, 2012). Using statistical models to capture and report these differences through the use of a feedback tool may help clinicians attend to important interactions that might otherwise be missed.

While some combinations may be easy to intuit (e.g., the interaction between depression and substance use), others could be more complex—involving three or more different constructs that may offer pertinent clues to the nature of a particular case. In the words of Andreasen and Grove (1982), "combinations of features

Samuel S. Nordberg, Louis G. Castonguay, and Andrew A. McAleavey, Department of Psychology, Pennsylvania State University; Benjamin D. Locke, Center for Assessment and Psychological Services, Pennsylvania State University; Jeffrey A. Hayes, Department of Educational Psychology, Counseling, and Special Education, Pennsylvania State University.

Samuel S. Nordberg is now at the Center for Healthcare Organization and Implementation Research in Boston, Massachusetts.

Correspondence concerning this article should be addressed to Samuel S. Nordberg, Center for Healthcare Organization and Implementation Research, 150 South Huntington Avenue, Boston, MA 02130. E-mail: samnordberg@gmail.com

too complex to grasp intuitively may yield better classifications than simple combinations" (p. 45). Identifying such complex interactions would require substantial training and time for a clinician, and would likely not fit in the course of routine care. Fortunately, the use of data-driven methods on large datasets holds some promise for uncovering meaningful multidimensional patterns within patient populations that can then be fed back to clinicians through computerized feedback tools. For example, latent profile analysis (LPA) has been used to model different response patterns on the interpersonal circumplex and has led to the identification of six groups of clients in psychotherapy with distinct interpersonal styles (Extraverted, Dominant, Arrogant, Cold, Submissive, and Unassuming), which in turn predicted outcomes in treatment for major depressive disorder (Cain et al., 2012). LPA has also been applied to scores on the Clinician-Administered Posttraumatic Stress Disorder (PTSD) Scale in order to identify subtypes of the disorder (Wolf, Lunney, et al., 2012; Wolf, Miller, Harrington, & Reardon, 2012). This work helped to establish a research basis for the new dissociative subtype of PTSD now enshrined in the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.) by identifying unique symptom profiles in a subset of veterans from various conflicts ranging from 15% to 32% of samples. Moreover, Latent profiles of patients with eating disorders have been developed as alternatives to diagnoses, and have been shown to predict mortality rates better than *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM–IV–TR*; American Psychiatric Association, 2000) diagnostics (Crow et al., 2012). It is also likely that meaningful clinical findings can be derived from LPA when applied to clients who are not categorized within one particular type of clinical problems (such as depression, PTSD, or eating disorders), as is it is the case in most naturalistic treatment settings. The current study is an attempt to use such an empirical approach to create multidimensional combinations of self-report symptoms that might provide helpful information to clinicians and administrators in the treatment and management of mental health problems for students in college counseling.

In the present study, client problems were measured by the College Counseling Assessment of Psychological Symptoms (CCAPS), which is a multidimensional assessment tool designed to help college counseling centers deliver patient-focused treatments to their students (Locke et al., 2011, 2012; McAleavey, Nordberg, Hayes, et al., 2012). The measure is currently supported by a suite of interpretation tools that provide feedback on subscale elevations, offer cut scores for assessing presence of particular disorders, and track change over time. In a recent poll of over 600 clinical staff members in counseling centers (including 96 directors), 388 suggestions for research focus on the CCAPS were given. The top two requests were for further research into using the CCAPS to better understand client characteristics and treatment outcomes (Center for Collegiate Mental Health, 2015).

Currently, CCAPS tools are based on single subscale averages drawn from very large sets of data. With the goal of better capturing clinically relevant differences between students, the CCAPS might be improved by disaggregating these large pools of students into qualitatively distinct subpopulations, characterized by multiple subscales. An optimal feedback system for the CCAPS would help guide clinicians' case formulation and treatment decisions by identifying important interactions between CCAPS subscales, reporting meaningful characteristics of a case for triage, and by predicting the likely response to and dose of treatment. For example, there might be significant differences between students with elevated Depression and Academic Distress if one group has elevated Family Distress and the other does not. This could, for example, indicate important challenges with support that make recovery in treatment more difficult. This might then be reflected in the amount of time spent in treatment, and the overall effect of that treatment.

Testing each of the possible combinations of CCAPS subscales would be extremely inefficient. Specifically, there are 256 possible combinations of the eight CCAPS subscales (and that is if we just consider them as elevated or not elevated). To identify which combinations of the CCAPS subscales best model students in counseling, we elected to use latent profile analysis (Muthén, 2001). As described in more details in the Method section, LPA was conducted as an attempt to model the variability between students by creating subgroups reporting similar symptoms across the CCAPS subscales.

Using large sets of data, as is the case in the present study, it is inevitable that a number of statistically distinct profiles will emerge, based on the currently available model-fitting statistics. Therefore, it has been recommended that when using methods that identify profiles, caution be taken in the interpretation of results (Bauer & Curran, 2003, 2004). In particular, it is important not to reify emergent profiles or to treat them as anything more than statistical means for parsing complicated data. To avoid these pitfalls, we followed the recommendations of Bauer and Curran (2003, 2004) by attempting to validate profiles of symptoms not by their statistical significance, but by their clinical significance—in this case, their ability to provide useful feedback to clinicians in routine care. Specifically, we set two criteria by which to evaluate profiles. These were directly related to clinical tasks that are present on a daily basis in college counseling and in behavioral health treatments more broadly. First, profiles must describe groups of patients with meaningful differences in their makeup—differences that would be useful to a clinician conducting an initial assessment with the purpose of triage to appropriate services. We broke this into two specific parts—predicting risk (of harming self or others) and guiding assessment (e.g., pointing to areas for further exploration, ruling out diagnoses, identifying areas of cultural importance). Second, profiles must facilitate the prediction of treatment outcome, including the dose of treatment and likelihood of success.

Method

Broadly, our analytical process tested the same analytical procedure (latent profile analysis) on two separate samples of CCAPS-62 assessments given to students seeking counseling services (which we refer to as Model 1 and Model 2). In order to assess the reliability of Model 1, the same exploratory LPA procedure was performed on a novel sample of identical size (Model 2). To determine the clinical usefulness of profiles, we examined the demographics and psychosocial indicators of each profile, with a particular focus on items related to risk. Following this, we examined whether profiles were useful for indicating diagnosis and, lastly, for predicting overall change during counseling. This process is detailed below.

Participants

The Center for Collegiate Mental Health (CCMH) is a practice–research infrastructure that currently includes more than 290 university and college counseling centers, each with independent review board approval for data sharing (Castonguay, Locke, & Hayes, 2011). We used two large samples drawn from archival datasets routinely collected by CCMH and accessible by permission of the organization. Demographics for all participants can be found in Table 1. CCAPS questionnaires were administered to students on a laptop or computer kiosk, before their first session, and as part of routine care.

Model 1. Participants for Model 1 (the exploratory model) were 19,247 college and university students initiating treatment at one of 52 CCMH centers participating in a large pilot study in the fall of 2008. This sample has been extensively described in Locke et al. (2011).

Model 2. Participants for Model 2 (the confirmatory model) were 19,247 students randomly selected from the CCMH data repository, from data collected between September 2010 and June 2011 at the first session of treatment.

Participants for diagnostic analyses. Data for analyses on diagnosis were 3,203 cases collected at the first session from two university counseling centers along with *DSM–IV–TR* diagnosis. This sample also included whether the first session for the client was a crisis intake.

Participants for outcomes analyses. Data for analyses on outcomes were 9,357 cases of repeated-measures data drawn from the CCMH data repository for the academic year from September 2010 through June 2011. In order to be included in the analyses, participants were required to have at least two measurement occasions during treatment, so that pre–post change could be evaluated.

Table 1
Demographics for All Participants

Variable	Model 1	Model 2	Diagnosis	Outcome
Age (*SD*)	22.6 (5.07)	21.3 (4.9)	22.7 (5.1)	23.1 (5.3)
Sex				
Male	35.0%	37.9%	42.3%	34.2%
Female	64.0%	61.9%	57.7%	65.6%
Transgender	1.0%	.2%	.0%	.2%
Ethnicity				
White	70.3%	71.7%	75.6%	72.1%
African American	7.6%	7.9%	6.2%	7.0%
Asian	6.3%	5.7%	6.3%	6.4%
Hispanic	5.8%	6.7%	5.4%	6.6%
Multiracial	3.2%	3.7%	2.4%	3.5%
Other	4.3%	.8%	.3%	
Did not respond	2.4%			
Sexual identity				
Straight	—	89.2%	90.7%	88.1%
Gay	—	2.2%	1.6%	2.5%
Lesbian	—	1.3%	.8%	1.9%
Bisexual	—	3.0%	1.7%	2.9%
Questioning	—	1.4%	.9%	1.9%
Did not respond	—	2.9%	4.3%	2.7%
Academic status				
Freshman	18.1%	19.6%	17.6%	18.1%
Sophomore	19.7%	19.1%	17.3%	17.9%
Junior	22.1%	23.3%	24.5%	22.9%
Senior	22.8%	22.0%	25.9%	22.3%
Graduate/other	17.3%	16.0%	14.7%	18.8%

Note. Graduate/other includes students who endorsed "professional studies" in addition to typical graduate work. *SD* = standard deviation.

Measures

The CCAPS-62 (Locke et al., 2011; McAleavey, Nordberg, Hayes, et al., 2012) is a 62-item self-report measure of psychological distress specifically tailored for use in college and university counseling centers. The instrument breaks out eight separate, factor-derived subscales: Depression, Generalized Anxiety, Social Anxiety, Academic Distress, Eating Concerns, Substance Use, Hostility, and Family Distress. Each subscale is built as an average of its respective questions, which are rated on a Likert-type scale ranging from 0 to 4 (*not at all like me* to *extremely like me*). The scale scores have excellent psychometric properties, including factor structure, convergent validity, and test–retest reliability (1-week test–retest ranged from r = .78 to 0.93 for the eight subscales).

The CCAPS-34 (Locke et al., 2012) is an abbreviated version of the CCAPS-62. It retains seven of the eight CCAPS-62 scales, dropping the Family Distress subscale and the Substance Use subscale from the CCAPS-62 and measures only Alcohol Use on the CCAPS-34. It was designed for use on a repeated basis, and has good internal consistency (α = 0.76–0.89) and 1- and 2-week reliability (r = 0.74–0.87). It has been shown to be sensitive to change (Youn et al., 2012).

The Standardized Data Set (SDS) collects information on demographics, school status, and past psychosocial history. For the purposes of Models 1 and 2, some SDS items were recoded. Specifically, responses of 1 = "never," 2 = "prior to college," 3 = "after starting college," and 4 = "both prior to and after starting college." These were recoded to 0 ("never") and 1 (all other responses, indicating "at some point").

Procedure

CCMH participating centers use the CCAPS-62 and −34, as well as the SDS, with every client at intake. The CCAPS-62 and SDS are typically given prior to their first appointment, and some participating centers administer the CCAPS-34 throughout treatment, or at termination.

Analytical Procedure

Latent class analysis (Muthén, 2001), or in the case of continuous variables, LPA, is a method used to identify hidden or latent subpopulations with highly similar features. Model building with LPA involved starting with the simplest model (a two-profile model) and testing whether two profiles fit the data better than a single profile. Then a three-profile solution would be compared to the two-profile solution, until model fit was no longer improved. We used the Akaike Information Criteria and the bootstrapping likelihood ratio test (BLRT; McLachlan & Peel, 2000) as our fit statistics. The BLRT in particular has been shown to most reliably reflect goodness of fit in simulations (Nylund, Asparouhov, & Muthen, 2007). To increase clinical usefulness, an additional criterion was set for constraining the analysis: If a model contained a profile representing less than 1% of the total sample, analyses would be discontinued. As mentioned in the introduction, it was expected that our analyses would result in a large number of statistically significant profiles, due to our very large datasets. We elected to follow the standard methods for testing for model significance, with the important caveat that significance in these models did not indicate clinical utility. Indeed, a nonnormal distribution for a homogeneous population could be bifurcated into many segments, each simply reflecting an artificial chunk of an otherwise continuous distribution (Bauer & Curran, 2003). This can be influenced by skew, for example, which could be reasonably expected in the distributions of several of our subscales (e.g., Substance Use or Eating Concerns). Only in testing the clinical usefulness of profiles would we truly begin to validate our findings.

In order to apply the results of the LPA to only those participants who are well classified by the model, individuals were excluded from the study if their highest posterior probability did not meet or exceed 0.70, or a 70% likelihood of being in their assigned profile. This procedure addresses an important challenge related to grouping individuals into a limited number of profiles. Inevitably, some individuals will be poorly grouped (i.e., they will look somewhat like people in one group, and somewhat like people in another group), due to the fuzzy edges of taxa based on continuous variables (Meehl, 1992) wherein certain individuals lie at the border between two groups. By selecting, a priori, a "good-enough" level of fit, we hoped to include in the final model only those individuals well described by it. We elected this novel procedure because our models are intended for practical use, rather than description of a sample. Thus, we did not wish to include as members of one particular profile those students who were largely described by two or more. For example, a student might have a posterior probability of 60% likelihood of being in Profile 1 and 40% likelihood in Profile 2. We determined that this was not a clear-enough distinction for clinical use, and therefore set our threshold at 70%. This allowed us to focus on the clinical usefulness of profiles for students who are, based on our criteria, adequately described by them.

We determined that a rational-empirical approach would be best suited to assessing whether or not the emergent profiles replicated. Profiles from Model 1 and Model 2 were represented graphically and similar profiles were identified via visual inspection. This visual inspection was further informed by comparison of the relative size (i.e., proportion of the total sample) of each profile with its visually identified counterpart. These rationally derived pairs were then empirically tested for statistical similarity.

An exploratory multivariate analysis of variance conducted on the subscales of the two full samples indicated small but significant differences between samples on the Generalized Anxiety, Family Distress, Substance Use, and Hostility subscales at $p < .001$.[1] Given the size of the samples, power to detect extremely small and possibly clinically meaningless differences was substantial for all but the smallest of groups. Therefore, we standardized each subscale score based on the sample means and standard deviations, and used Cohen's d (Cohen, 1988) as an estimate of the size of the mean differences between paired groups on each of the eight subscales. If paired profiles had subscale differences of a moderate or large effect size, a rationale based on clinical indicators had to be provided for the profile to be considered a match. If none such could be provided, the profile would be considered unmatched.

Diagnosis. The parameter estimates from Model 2 were used to force the diagnostic data into a model with similar profile means and standard deviations. Students with posterior probability less than 0.7 were excluded from the final model. Diagnoses were broken down in to 7 categories:[2] Depressive Disorders included diagnoses of major depressive disorder and dysthymia; Anxiety Disorders included diagnoses of generalized anxiety disorder, panic disorder, and social anxiety disorder. Anorexia, bulimia, and eating disorder not otherwise specified were kept separated because profiles exhibited substantially different incidence of each, and they appeared to be indicative of markedly different phenomena; however, we also collapsed these into a single category in order to test the overall hypothesis regarding profiles labeled with "Eating Concerns." Alcohol dependence and abuse were combined into one category. Personality disorders were of such low base rates that they were aggregated into one category, capturing any diagnosis on Axis II. Schizophrenia was maintained as its own diagnostic category. Lastly, because the academic distress subscale ties directly to concern about academic performance, we included the V-code "Academic Problems." Chi-square tests of frequency were applied to the profiles for each diagnostic category, to examine whether differences in the incidence of diagnoses were statistically significant.

Outcomes. We identified three outcomes of interest: engagement in treatment, treatment length, and pre–post change. For the first, we used available session-data from the confirmatory LPA ($n = 16{,}684$ of the 19,247 total) to look at the proportion of each profile who attended only one session of treatment, as well as the ratio of cancelled or no-showed appointments to attended appointments. For treatment length, we compared profiles with pre–post data ($n = 4{,}294$) on the average number of sessions using an analysis of variance (ANOVA).[3] Pre–post change was calculated using the first and last CCAPS administered. This is predicated on the assumption that the last CCAPS is on or close to the end of treatment date.[4] These analyses closely reflect the way in which the interpretation tools for the CCAPS operate—calculating pre–post differences from intake to the most recent assessment, and thus are readily transferrable to existing tools.

Consistent with methods used by Kraus and colleagues (2011), we elected to use two forms of analysis to assess differences in

[1] Note that these differences were very small—on the order of hundredths—with the largest difference no greater than 0.05. That such a small difference should be so significant is an indication of the power associated with the very large samples used in this study.

[2] The selection of diagnostic categories was based on examination of the results with all diagnoses. Some diagnostic groups had extremely low count (e.g., severe Major Depressive Disorder, borderline Personality Disorder [PD]), and seemed unreliable as indicators of prevalence. These groups were combined into larger groups (e.g., Depression, PD) for the purposes of these analyses.

[3] We elected to include those students who were likely in either weekly or biweekly treatment. Therefore, the number of sessions had a higher concordance with overall dose of treatment than did total time in treatment.

[4] Centers that administer repeated CCAPS tend to administer a second CCAPS at, or close to, termination. Centers that administer the CCAPS more frequently will do so at an interval, such as every three sessions, or every first week of the month. The current data do not assess for the methods used by counseling centers, and thus it is impossible to know whether the last CCAPS was administered at the very last session. However, given that the CCAPS is used for program evaluation and performance benchmarking, it seems fair to assume that centers administer the "post" CCAPS very close to the final session, in order to best capture the effect of their interventions.

change over time. First, we used the clinically significant change formulae developed by Jacobson and Truax (1991) as a measure of successful recovery. In addition, we calculated effect size (Cohen's *d*) differences from pre to post for all students in each profile, in order to provide an estimate of overall change.

We chose to focus our analyses of change on four CCAPS-34 subscales: Depression, Alcohol Use, Eating Concerns, and Hostility. Depression was selected, as it has been shown be an adequate proxy for overall client distress (Nordberg et al., 2016) and appears to mimic the profile differences in the Generalized Anxiety, Social Anxiety, and Academic Distress subscales (we observed that these scales were roughly equivalent to one another for all profiles). Alcohol Use, Eating Concerns, and Hostility were selected for their unique contribution to the formation of profiles, and because they have been shown to capture client symptoms that are different from distress. We could not examine change on the Family Distress subscale, as it is not included on the CCAPS-34.

Based on the findings from the two exploratory LPAs, CCAPS-62 subscale scores from the first session for each student with two or more repeated measures ($N = 9{,}357$) were first fit to a similar model by forcing the model to converge to the profile means and standard deviations from the confirmatory model. Subjects with less than 70% posterior probability of having been correctly assigned to a profile ($N = 3{,}118$, or 33.3%) were removed from the data set. The remaining subjects ($N = 6{,}239$), were retained for longitudinal examination of change over time. In order to exclude cases that were not likely continuous treatment courses of psychotherapy we calculated the ratio of time in days between the first and last CCAPS administration and the total number of attended sessions and only included data for clients with a ratio of 1 session per 17 days in treatment or better (designed to capture weekly and biweekly treatment modalities). This resulted in the elimination of an additional 1,945 cases, and a final pre–post dataset of 4,294 cases.[5]

The CCAPS-62 and −34 have slightly different subscale means, and cannot be directly compared to one another. Thus, in our analyses of pre–post change, we scored CCAPS-62s as −34s, since all of the CCAPS-34 items appear on the CCAPS-62, and there were more CCAPS-34 administrations at posttreatment.

Results

LPA Statistical Significance and Model Characteristics

Model 1. Based on the model-fit criteria (see Table S1 in the online supplemental materials), a 17-profile model was a significant improvement over models with fewer profiles. Although an 18-profile model was an even better fit according to fit indices, one profile included less than 1% of the total sample size. Thus, given our established constraint, we selected the 17-profile model. Poorly classified participants were removed from the model if they did not have a posterior probability of 0.70 or greater. These individuals represented a substantial portion of the overall sample ($N = 5{,}785$, or 30.1%). Means and relative sizes of the profiles can be seen in Table 2. Demographic and psychosocial information can be found in Tables S3 and S6 in the online supplemental materials.

Model 2. Based on the model-fit criteria, a 17-profile model was selected as the best-fit to the data. As was the case in Model 1, an 18-profile model significantly improved the fit to the data, but one of the emergent profiles accounted for less than 1% of the total participants. Based on posterior probabilities, 6,052 individuals (31.4% of the sample) were removed from the model because their posterior probabilities fell below the 0.70 threshold established in Model 1. This number is comparable with those removed in Model 1 ($N = 5{,}785$, 30.1%). Means and profile sizes for the 17 profiles can be found in Table 2. Demographics and psychosocial information can be found in Tables S4 and S6 (online supplemental materials).

Profile matching. The rational matching approach resulted in 16 profiles that were considered promising replicas of one another (see Table 2). The examination of effect size differences (see Table S2, online supplemental materials) between paired profiles confirmed the rejection of one profile (Profile 12) and required further rational examination of three (6, 11, and 16). In most cases, effect size differences between pairs were small ($d < 0.20$), and supported matches made by visual inspection. Profiles 6, 11, and 16 had one medium effect size ($0.50 < d < 0.80$) difference on one subscale (Substance Use, Hostility, and Hostility, respectively). The rationale to keep these last three is described following the results for demographic and psychosocial matching. An examination of the relative sizes of each replicated profile (found in Table 2) indicates that the replicated profiles were similar in the proportion of overall participants captured.

Demographic matching. Tables 3S and 4S (online supplemental materials) show the gender breakout for each replicated profile, as well as the base rates for gender information. We examined differences between profiles from each matched pair (e.g., Profile 1 from Models 1 and 2). Analyses of effect size differences between matched profiles indicated that for all replicated profiles, no gender differences approached even a small effect size, indicating stability in the gender proportions for each profile across Model 1 and Model 2.[6]

Tables 3S and 4S (online supplemental materials) also show the ethnic breakout for each replicated profile. Analysis of effect-size differences within replicated profiles indicated that there were no differences of even a small size for matched profiles from Model 1 and Model 2. It can be difficult to interpret differences between low-base-rate groups, and these results should be accepted with caution. As a check on the influence of low-base-rate groups, we reran our effect-size analyses using a White/non-White variable, and replicated the results above. Some differences between profiles (e.g., between Profile 1 and Profile 5) were quite significant. For example, Profiles 5, 8, and 16 consistently demonstrated substantial overrepresentation of non-Whites.

Psychosocial matching. We examined psychosocial history variables contained in the SDS, which are displayed in Table 6S (online supplemental materials). We generated estimates of effect-size differences between samples for each profile using the phi coefficient for dichotomous data, and eta-squared statistic for

[5] This final dataset consisted of 2,632 students attending centers without session limits, and 1,519 attending centers with reported session limits (there were 143 missing responses). An ANOVA of these groups (including those missing responses) indicated no mean differences in overall treatment length, $F(1, 4294) = 1.889$, $p = .170$, and thus we included all students in our analyses.

[6] Results were excluded for the sake of space, and are available on request from Samuel S. Nordberg.

Table 2
Side-by-Side Comparison of Standardized Subscale Means and Profile Size for Replicated Profiles

Profile	Subscale means								Profile size
	DEPRS	GANX	SAND	ACDIS	FAMILY	EAT	SU	HOST	
1	−1.10	−.93	−.71	−.75	−.58	−.51	.81	−.57	5.34%
1	−1.00	−.83	−.69	−.73	−.51	−.48	.87	−.57	4.75%
2	−.63	−.50	−.26	−.59	−.22	−.36	2.12	.00	2.52%
2	−.75	−.65	−.37	−.70	−.36	−.43	2.24	−.35	1.95%
4	.23	.22	.27	.09	−.01	−.31	.87	.08	8.12%
4	.27	.30	.39	.20	.01	−.15	.94	−.02	7.20%
11	.88	.68	.67	.28	.44	−.23	2.41	.91	3.45%
11	.73	.68	.66	.42	.27	.00	2.55	.47	2.96%
3	.55	.48	.31	.57	.36	1.69	−.53	.12	6.08%
3	.85	.64	.51	.80	.44	1.74	−.49	.18	6.63%
9	−.63	−.55	−.68	−.29	−.33	1.61	−.52	−.60	3.33%
9	−.48	−.46	−.60	−.32	−.22	1.66	−.33	−.47	4.25%
6	−.23	−.21	−.11	−.16	−.11	1.69	1.00	−.14	1.99%
6	.01	.00	.00	−.10	−.17	1.98	1.65	−.21	1.92%
13	1.02	.96	.66	.72	.70	1.54	.95	.70	3.34%
13	1.20	1.07	.87	.85	.66	1.81	1.19	.67	2.76%
7	−.33	−.17	−.10	−.12	−.15	−.47	−.61	−.41	19.21%
7	−.28	−.18	−.09	−.05	−.07	−.45	−.58	−.4	22.14%
10	−1.24	−1.10	−.99	−.90	−.70	−.66	−.61	−.85	18.98%
10	−1.23	−1.10	−.97	−.91	−.67	−.66	−.57	−.83	19.66%
5	.01	.06	−.14	−.25	.87	−.30	−.53	1.71	2.68%
5	.09	.13	−.22	−.12	.57	−.26	−.44	1.60	3.53%
8	1.45	1.20	.89	.81	1.00	−.11	−.57	1.64	4.34%
8	1.40	1.18	.87	.87	1.10	−.13	−.52	1.68	4.70%
14	1.34	1.10	.91	.77	.80	−.28	.99	1.46	2.96%
14	1.31	1.27	.86	.67	.74	−.15	1.17	1.65	2.83%
15	.85	.55	.64	.69	.06	−.37	−.59	−.13	11.78%
15	.94	.65	.77	.68	−.01	−.40	−.56	−.11	10.15%
16	1.57	1.46	1.04	1.21	1.03	1.88	−.44	1.53	2.78%
16	1.65	1.49	1.06	1.03	1.19	1.83	−.14	2.25	1.91%
17	1.52	1.50	1.10	.83	1.17	1.70	2.28	1.98	1.75%
17	1.50	1.51	1.02	.74	1.15	1.51	2.71	1.98	1.69%

Note. Exploratory means and sizes are listed above respective confirmatory means and sizes. DEPRS = Depression subscale; GANX = Generalized Anxiety subscale; SANX = Social Anxiety subscale; ACDIS = Academic Distress subscale; FAMILY = Family Distress subscale; EAT = Eating Concerns subscale; SU = Substance Abuse subscale; HOST = Hostility subscale; Profile size = the proportion of well-fit subjects belonging to a particular profile.

continuous data. Fewer than 5% (16 out of 352) of the tests indicated a small or larger effect-size difference between matched profiles, suggesting that psychosocial patterns were largely replicated across Model 1 and Model 2. These results should be taken with caution, however, because low base-rate events such as hospitalization can cause tests of correlation to underestimate the amount of covariation.

We determined that all 16 visually matched profiles, including the three profiles with moderate effect-size differences (6, 11, and 16) were reasonable replications of one another, and indicated that our statistical model was quite stable. We based this decision on the examination of distress patterns, profile size, and demographics of each matched pair.

Diagnosis results. Posterior probabilities were used to identify individuals who were poorly classified (N = 1,143; 35.7%), leaving a sample size of 2,060 for further analysis. Table 3 illustrates the base rates and frequencies of particular diagnoses for each of the 16 stable profiles. Table 5S (online supplemental materials) shows the results of omnibus chi-square tests for each diagnostic category. Except for the V-code for Academic Problems, all of the tests were significant at $p < .001$, though there were substantial differences in the size of the effects, in terms of the variation explained.

Outcomes results. Using posterior probabilities of 70% or greater, we created a final dataset of 4,294 students who could be classified into profiles. The results of an omnibus ANOVA revealed significant differences between profiles in number of sessions attended, $F(16, 4{,}293) = 5.12$, $p < .001$, with 94 missing responses. Figure 1S (online supplemental materials) illustrates these differences graphically, with 95% confidence intervals. The average number of sessions ranged from 7.34 for Profile 1 to 13.03 for Profile 17.

Missed opportunities (the ratio of cancellations and no-shows to attended appointments), failure to engage (the proportion of students attending only one session), effect sizes, cut-score statistics, and overall clinically significant change indicators are shown in Table 4. The percentage of clients meeting the criteria for clinically significant change is given as the percentage of the subset of clients whose pretreatment scores were above the clinical cut-off. Missed opportunities ratios ranged from 23.36% to 42.16%, and an omnibus ANOVA found significant differences between profiles, $F(16, 15{,}720) = 7.605$, $p < .001$, with 964 missing responses.

Table 3
Diagnostic Group by Profile

		Primary Substance Use				Primary Eating		Eating and SU		Undifferentiated		Family and Hostility					
Diagnosis	Base	1	2	4	11	3	9	6	13	7	10	5	8	14	15	16	17
Depression	26.00%	5.30%	9.70%	27.10%	48.90%	45.20%	7.10%	17.90%	50.00%	12.20%	3.50%	23.10%	62.20%	66.00%	42.80%	61.90%	71.00%
Anxiety	29.60%	14.40%	19.40%	34.50%	30.40%	30.80%	23.80%	26.80%	37.30%	34.10%	19.00%	23.10%	29.70%	40.00%	38.50%	45.20%	29.00%
GAD	18.00%	7.50%	10.80%	21.20%	21.70%	19.20%	14.30%	10.70%	25.50%	21.00%	8.10%	11.50%	23.00%	28.00%	25.80%	23.80%	16.10%
Social phobia	6.40%	1.10%	4.30%	8.40%	9.80%	4.80%	2.40%	1.80%	7.30%	7.40%	3.20%	3.80%	12.20%	6.00%	10.70%	4.80%	16.10%
Any ED	7.40%	1.10%	1.10%	.50%	1.10%	38.50%	57.10%	53.60%	13.60%	2.40%	2.30%	.00%	2.70%	.00%	1.70%	26.20%	6.50%
Anorexia	1.30%	.00%	.00%	.50%	.00%	9.60%	4.80%	8.90%	1.80%	.20%	.60%	.00%	.00%	.00%	.70%	7.10%	.00%
Bulimia	2.00%	.50%	.00%	.00%	.00%	8.70%	21.40%	23.20%	3.60%	.00%	.00%	.00%	.00%	.00%	.00%	9.50%	.00%
ED NOS	4.50%	.50%	1.10%	.00%	1.10%	21.20%	38.10%	25.00%	10.00%	2.10%	1.60%	.00%	2.70%	.00%	1.30%	9.50%	6.50%
Alcohol Dx	13.70%	41.70%	74.20%	13.80%	41.30%	1.00%	2.40%	26.80%	7.30%	1.20%	7.40%	.00%	.00%	14.00%	1.30%	.00%	25.80%
Academic Dx	6.90%	3.70%	5.40%	7.90%	4.30%	4.80%	2.40%	1.80%	8.20%	9.80%	7.10%	3.80%	9.50%	8.00%	8.00%	4.80%	3.20%
PD	4.70%	1.60%	3.20%	3.00%	7.60%	4.80%	.00%	5.40%	5.50%	4.10%	1.60%	3.80%	12.20%	8.00%	5.70%	14.30%	25.80%
Schiz	.50%	.00%	.00%	.50%	1.10%	.00%	.00%	.00%	.00%	.20%	.30%	.00%	.00%	6.00%	.30%	2.40%	6.50%
Crisis	29.60%	9.10%	18.30%	31.00%	44.60%	24.00%	9.50%	17.90%	32.70%	27.40%	15.20%	38.50%	60.80%	54.00%	41.50%	64.30%	71.00%
Male	42.30%	62.40%	66.70%	57.20%	62.00%	16.30%	7.10%	10.90%	25.70%	36.00%	49.00%	50.00%	39.20%	52.00%	38.70%	23.80%	38.70%
Female	57.70%	37.60%	33.30%	42.80%	38.00%	83.70%	92.90%	89.10%	74.30%	64.00%	51.00%	50.00%	60.80%	48.00%	61.30%	76.20%	61.30%
C size	NA	7.76%	3.69%	9.77%	3.61%	5.20%	2.32%	2.96%	4.93%	18.59%	12.87%	2.24%	3.96%	2.69%	14.38%	2.15%	1.34%

Note. Profile 12 is not included in this table as it was eliminated due to poor match. Percentages in this table are not given relative to the base rate, but as the proportion of the sample with a particular diagnosis. SU = Substance Abuse subscale; GAD = generalized anxiety disorder; ED NOS = eating disorder not otherwise specified; Any ED = any eating disorder, made up of anorexia, bulimia, and ED NOS; Alcohol Dx = diagnosis of alcohol abuse or alcohol dependence; Academic Dx = V-code diagnosis of academic problems; PD = any Axis II diagnosis; Schiz = schizophrenia diagnosis; Crisis = the student's intake was an unscheduled crisis walk-in; C size = the size of the profile relative to the entire sample.

Failure to engage ranged from 15.22% to 28.43%, and an omnibus ANOVA indicated significant differences between profiles, $F(16, 16{,}667) = 8.700$, $p < .001$, with 17 missing responses. The specific effect of class membership on missed opportunities and failure to engage is described in greater detail in the Discussion section.

In general, clinically significant change statistics were difficult to use to compare profiles. In many cases, there was limited correspondence between treatment effect size and the proportion of subjects who met the criteria for clinically significant change. Because reliable change requires crossing the clinical cut-off, profiles with mean scores closer to the cut score appeared to evidence more clinically significant change than other profiles with nearly twice the effect size in pre–post change. For example, with regard to change in the Alcohol Abuse subscale, Profiles 14 and 17 had very similar proportions of members meet the criteria for clinically significant change (36.4% and 38.9%, respectively), while evidencing markedly different effect sizes ($d = -0.82$ and $d = -1.55$, respectively). Therefore, we determined to focus our interpretation on effect size differences between classes, as well as posttreatment functioning.

Next, we tested the hypothesis that there would be differences in treatment outcomes for clients in different classes. Because profiles varied widely in the initial severity of symptoms on each subscale, differences in outcome could simply be a result of these differences in severity. We examined pre–post change on the Depression, Eating Concerns, Substance Use, and Hostility subscales (as discussed above in the Method section), these appeared to offer the most unique contribution to differences (and similarities) between classes. Additionally, we only selected profiles with mean elevations above 1 standard deviation, since many profiles had almost a complete absence of elevation on certain subscales, and would not be relevant for comparison on outcomes. Because these groups still had small variations in starting severity, we included initial severity in our analyses.

Differences in outcomes between profiles are described in greater detail in the Discussion section. Omnibus analysis of covariance, accounting for pretreatment severity, indicated meaningful differences in total change between classes (3, 6, 9, 13, 16, and 17) on the Eating Concerns subscale, $F(5, 935) = 2.59$, $p = .02$. A significant effect of class was found for the Substance Use subscale (classes 2, 6, 11, 14, and 17), $F(4, 525) = 3.154$, $p = .01$. On the Depression subscale (for classes 8, 13, 14, 16, and 17), the effect of class on total change approached significance, $F(4, 784) = 2.211$, $p = .07$. There was no effect of class membership on pre–post change on the Hostility subscale, $F(4, 751) = 0.362$, $p = .84$.

Collapsing profiles. To facilitate interpretability, and as a further attempt to acknowledge the concerns raised by Bauer and Curran (2004) regarding the risk of conflating differences in kind with differences in degree, we rationally grouped profiles which appeared to reflect the same underlying problems at differing levels of severity. This process reduced the number of discrete profiles from 16 to 9 and simplified our clinical feedback. Our rationale for grouping was first to examine the relative subscale elevations, and to attempt to group profiles based on similar elevations on the Family Concerns, Eating Concerns, Substance Use, and Hostility subscales, while allowing the mood scales (Depression, Generalized Anxiety, Academic Distress, and Social

Table 4
RCIs, Cutoffs, and Clinically Significant Change

Variable	Base rate	Primary Substance Use				Primary Eating		Eating and SU		Undifferentiated		Family and Hostility					
		1	2	4	11	3	9	6	13	7	10	5	8	14	15	16	17
Treatment Sessions	9.42	7.34	7.50	9.44	8.93	9.79	9.55	9.85	11.05	9.40	7.57	9.19	9.85	10.33	9.71	10.66	13.03
Missed opportunities ratio	30.74%	25.37%	32.60%	35.30%	35.11%	31.92%	34.03%	33.57%	37.27%	29.38%	23.36%	42.16%	35.36%	37.38%	28.68%	32.78%	42.06%
% fail to engage	21.30%	28.16%	20.85%	20.27%	21.36%	15.22%	20.70%	19.91%	16.42%	20.49%	28.43%	21.25%	17.71%	18.70%	17.33%	16.67%	24.50%
Pre Dep Above cutoff	51.10%	.00%	2.80%	57.3%	89.00%	76.3%	5.00%	26.1%	97.60%	9.70%	.00%	43.1%	100%	98.80%	95.00%	98.90%	100%
Post Dep above cutoff	27.40%	1.80%	11.10%	24.40%	43.90%	33.90%	8.30%	10.90%	51.80%	10.10%	2.60%	26.20%	55.70%	51.80%	43.70%	65.90%	47.20%
Dep CSI		.00%	.00%	27.70%	30.10%	31.00%	33.30%	35.70%	34.60%	40.00%	.00%	35.70%	39.70%	40.50%	37.60%	32.20%	47.20%
Dep effect sizes	−.60	−.02	−.08	−.60	−.76	−.67	−.13	−.20	−.81	−.31	.07	−.37	−1.18	−1.16	−.93	−1.02	−1.32
Pre AU above cutoff	21.10%	69.60%	100%	59.80%	100%	.00%	.00%	82.60%	72.30%	.00%	.00%	.00%	.00%	77.60%	.00%	.00%	100%
Post AU above cutoff	14.30%	39.30%	66.70%	34.80%	68.30%	4.30%	.00%	52.20%	31.30%	2.70%	1.10%	1.50%	3.80%	31.80%	3.20%	8.00%	61.10%
AU CSI		23.10%	27.80%	17.30%	31.70%	.00%	.00%	21.10%	25.00%	.00%	.00%	.00%	.00%	36.40%	.00%	.00%	38.90%
AU effect sizes	−.23	−.56	−1.00	−.54	−1.31	.07	−.04	−.56	−.48	.03	.03	.10	.10	−.82	.05	.01	−1.55
Pre Eat above cutoff	21.80%	.00%	5.60%	.00%	.00%	95.70%	95.00%	95.70%	86.70%	.00%	.00%	4.60%	.00%	.00%	.00%	97.70%	86.10%
Post Eat above cutoff	18.50%	1.80%	8.30%	4.30%	7.30%	64.00%	43.30%	73.90%	63.90%	5.80%	1.50%	6.20%	3.80%	8.20%	8.80%	65.90%	47.20%
Eat CSI		.00%	.00%	.00%	.00%	17.40%	28.10%	15.90%	12.50%	.00%	.00%	33.30%	.00%	.00%	.00%	23.30%	22.60%
Eat effect sizes	−.16	.04	−.01	−.03	−.07	−.61	−.64	−.35	−.40	.07	.05	−.04	−.03	−.11	.01	−.74	−.67
Pre Hostile above cut	48.30%	23.20%	38.90%	54.90%	80.50%	51.10%	16.70%	47.80%	84.30%	26.90%	5.90%	100%	99.20%	96.50%	43.70%	94.30%	100%
Post Hostile above cut	36.30%	19.60%	27.80%	43.30%	57.30%	36.60%	13.30%	37.00%	54.20%	20.90%	5.60%	73.80%	71.80%	68.20%	33.30%	75.00%	86.10%
Hostility CSI		.00%	7.10%	10.00%	13.60%	14.70%	.00%	13.60%	20.00%	5.80%	.00%	23.10%	24.60%	20.70%	10.10%	14.50%	13.90%
Hostility effect sizes	−.37	−.04	−.16	−.26	−.61	−.28	.01	−.14	−.53	−.09	.03	−1.20	−1.09	−.98	−.11	−1.00	−1.21

Note. RCIs = Reliable Change Index; SU = Substance Abuse subscale; Treatment sessions = the average number of attended sessions for students with pre–post data; Missed opportunities ratio = the sum of cancelled and no-showed appointments divided by attended appointments; Fail to engage = the percentage of students who attended only one appointment; Dep = the Depression subscale; CSI = clinically significant indicator = the percentage of clients who began above the clinical cutoff, crossed below the clinical cutoff, and exceeded the RCI; AU = the Alcohol Use subscale; Eat = the Eating Concerns subscale; Hostile = the Hostility subscale.

Anxiety) to vary. This, we believed, was the best means to reflect differences in the degree of severity within a profile group, while also capturing differences in the kind of distress between profile groups.

Profiles with primary eating concerns we grouped into two types. Primary Eating Concerns (see Figure 1) contains two profiles with similar elevations in eating concerns and varying levels of mood symptoms. Eating Concerns and Substance Use (see Figure 1) are made up of two profiles with similar levels of both eating concerns and substance use, and varying degrees of mood symptoms. We collapsed all four profiles with primary substance use into one type (Primary Substance Use; Figure 2) characterized by varying levels of substance use, combined with varying levels of disordered mood.

Two profiles (Family and Hostility; Figure 3) captured elevations in the Family Concerns and Hostility subscales, at varying levels of elevated mood. Two undifferentiated profiles were grouped into one type at varying levels of severity (undifferentiated; Figure 4). Lastly, four profiles were too distinct for us to rationally group them, and we left these ungrouped (see Figure 5).

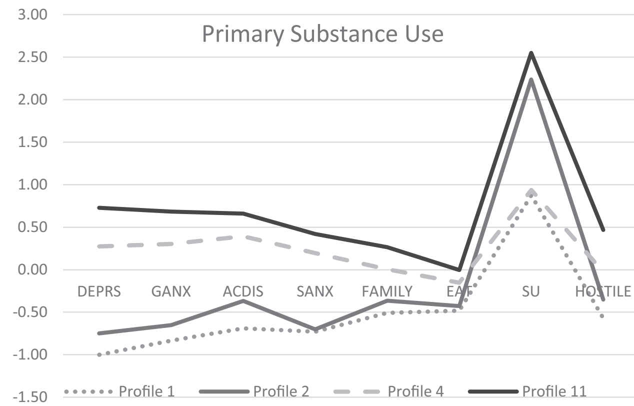

Figure 2. Primary substance use profiles. DEPRS = Depression subscale; GANX = Generalized Anxiety subscale; ACDIS = Academic Distress subscale; SANX = Social Anxiety subscale; FAMILY = Family Distress subscale; EAT = Eating Concerns subscale; SU = Substance Use subscale; HOSTILE = Hostility subscale.

Discussion

The goal of this study was to explore the possibility of discovering useful subpopulations of clients seeking counseling in routine care, and to test the reliability and clinical utility of those emergent profiles. Latent profile analysis resulted in two models with very high overlap, indicating that this method appears to produce reliable results across samples. Our analyses show that a division of large samples into profiles improved the model fit over one described by a single set of means, supporting the statistical significance of the selected models. It is important to note here that, although 17 profiles emerged from our analyses, there is no reason to believe that this is the "right" or even optimal number of profiles. Our methodological criteria, more than anything else, determined this number, and we encourage readers to focus instead on whether the emergent profiles are useful. Evaluation of the clinical utility of the profiles indicates broadly that profiles can be of use for both alerting clinicians to important characteristics of particular students and for predicting the duration and success of treatment. In the current study, 16 statistically reliable profiles appeared to reflect 9 clinically distinct kinds of problem areas. Given the wide use of the CCAPS (with over 100,000 unique students from 2013–2014), these profiles, or profile groups, have the potential to enhance assessment and outcome monitoring in the routine delivery of counseling services. For clinicians who may have only a few minutes to review CCAPS results before starting a session, such feedback can be valuable if presented succinctly in a feedback report.

Profiles appeared to capture meaningful differences between groups of students on important psychosocial variables such as history of suicide, violence, and self-harming behaviors. The proportion of individuals endorsing a history of self-harming behaviors, for example, varied from 4.9% to 59.7% across the profiles. Similarly, for prior suicide attempt, the proportions ranged from 2% of a profile to 22.1%. These findings can provide a unique and

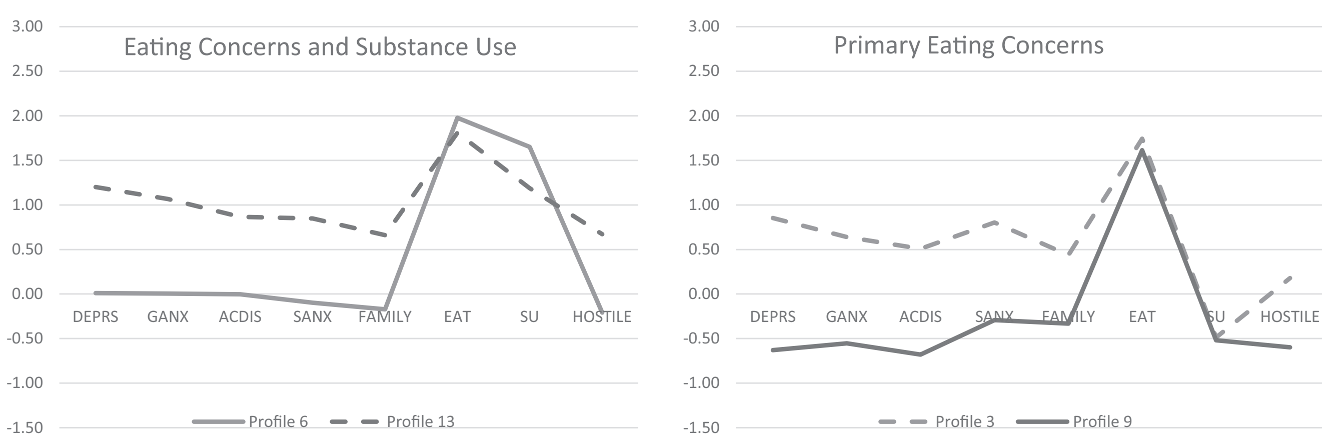

Figure 1. Eating concerns profiles. DEPRS = Depression subscale; GANX = Generalized Anxiety subscale; ACDIS = Academic Distress subscale; SANX = Social Anxiety subscale; FAMILY = Family Distress subscale; EAT = Eating Concerns subscale; SU = Substance Use subscale; HOSTILE = Hostility subscale.

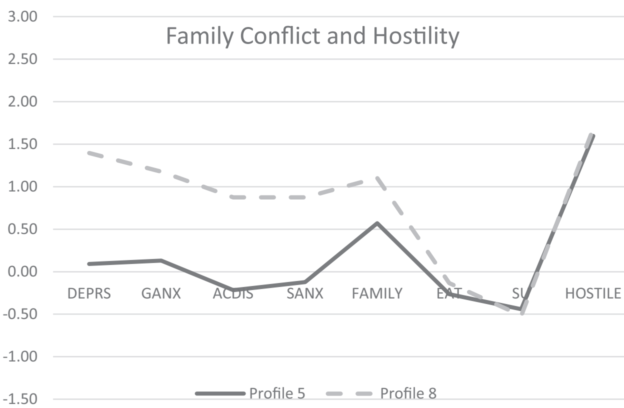

Figure 3. Family conflict and hostility profiles. DEPRS = Depression subscale; GANX = Generalized Anxiety subscale; ACDIS = Academic Distress subscale; SANX = Social Anxiety subscale; FAMILY = Family Distress subscale; EAT = Eating Concerns subscale; SU = Substance Use subscale; HOSTILE = Hostility subscale.

helpful feedback to clinician: They suggest focusing initial assessment sessions on evaluating the most likely risks, even if the particular student has not endorsed them. There are many reasons why a particular student might not endorse risk items, such as stigma or fear that they might be forced to withdraw from classes. Providing clinicians with the knowledge that clients reporting similar symptoms endorsed high-risk items can facilitate a meaningful conversation that could begin with, “Many of your peers who report similar challenges also report that they have harmed themselves.” This could allow a clinician to gently broach the subject while also conveying the sense that the student would not be alone in this behavior if it is present.

In addition, profiles indicated differences between students with respect to the duration and likely success of treatment. While approximately 40% of profiled patients in our samples ended treatment after 7 sessions, roughly 10% averaged around 12 or 13 sessions. It is worth noticing that in 2014, 51 (37.2%) of 140 colleges and universities that provided data to CCMH had session limits (Center for Collegiate Mental Health, 2015). Of these, roughly one-third had session limits of 10 or fewer. For centers that have adopted (or have been forced to abide with) blanket policies on session limits, our findings can raise important clinical and administrative implications. For example, a center could compellingly advocate to organization leadership that members of certain profiles require an extension of standard practice, and may justify hiring additional staff to ensure that effective treatment is delivered. With regard to engagement in treatment and attendance, results appear to indicate that roughly one in five students do not attend sessions after their first, and that they cancel or no-show roughly 3 appointments for every 10 they attend. Notably, with the exception of Profile 17, it appears that students in profiles with lower rates of engagement also have lower rates of missed opportunities. One conclusion is that some students with low distress (e.g., those in Profiles 1 and 10) are likely to self-select into or out of treatment after a first session and that, upon committing, are likely to stick with their decision.

Based on these initial results, it may be possible to predict when a therapist would have new availability in their caseload, based on the mix of patients fitting different profiles. For example, by understanding average treatment lengths, one could build into a schedule anticipated discharge dates and estimate new openings. Lastly, profiles could assist counseling centers in developing a better understanding of treatment needs and how to meet them. Here, for example, a center could examine the proportion of students fitting particular profiles for disordered eating and determine whether they wanted to develop expertise in those treatments in-house, or build partnerships with community experts instead. Profile data could augment conversations with senior leadership for additional resources or trainings by providing easily communicated and well-validated data.

Because the selected method resulted in a wide array of clinically relevant findings, we have broken the remainder of the discussion into several sections, each related to the particular grouping of the subscales noted in the Method section. We explicitly lay out a rationale for the use of profiles in clinical tools, including a more detailed description of which profiles appear to capture similar phenomena, and how the information can be provided to clinicians to augment their judgment. The first five sections are grouped by the prevailing symptom profile. The fifth, and last, section highlights profiles which appear to capture important cultural and ethnic implications.

Primary Eating Concerns (Profiles 3, 6, 9, and 13; Figure 1)

Four profiles were characterized by elevated eating concerns at varying levels of mood and substance use and, together, account for 14.3% of the total assigned sample. We argue that these profiles are best grouped into two subsets of two profiles each—Primary Eating Concerns and Eating Concerns and Substance Use—and presented in feedback as being differentiated by elevations in substance use and distress symptoms.

Assessment and treatment planning. The more complex interaction among distress symptoms, eating concerns, and substance use indicates several pieces of clinical feedback that could facilitate assessment and treatment planning. In both groups, in-

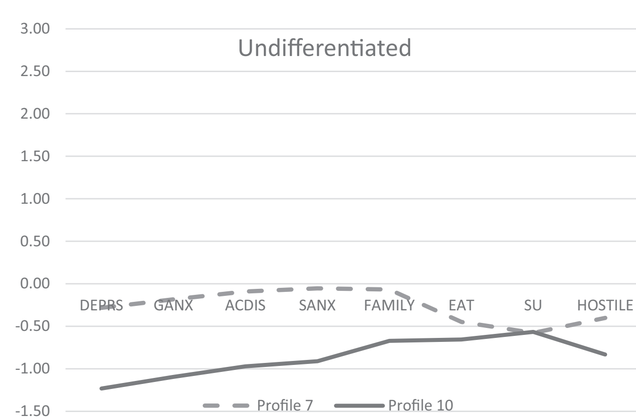

Figure 4. Undifferentiated profiles. DEPRS = Depression subscale; GANX = Generalized Anxiety subscale; ACDIS = Academic Distress subscale; SANX = Social Anxiety subscale; FAMILY = Family Distress subscale; EAT = Eating Concerns subscale; SU = Substance Use subscale; HOSTILE = Hostility subscale.

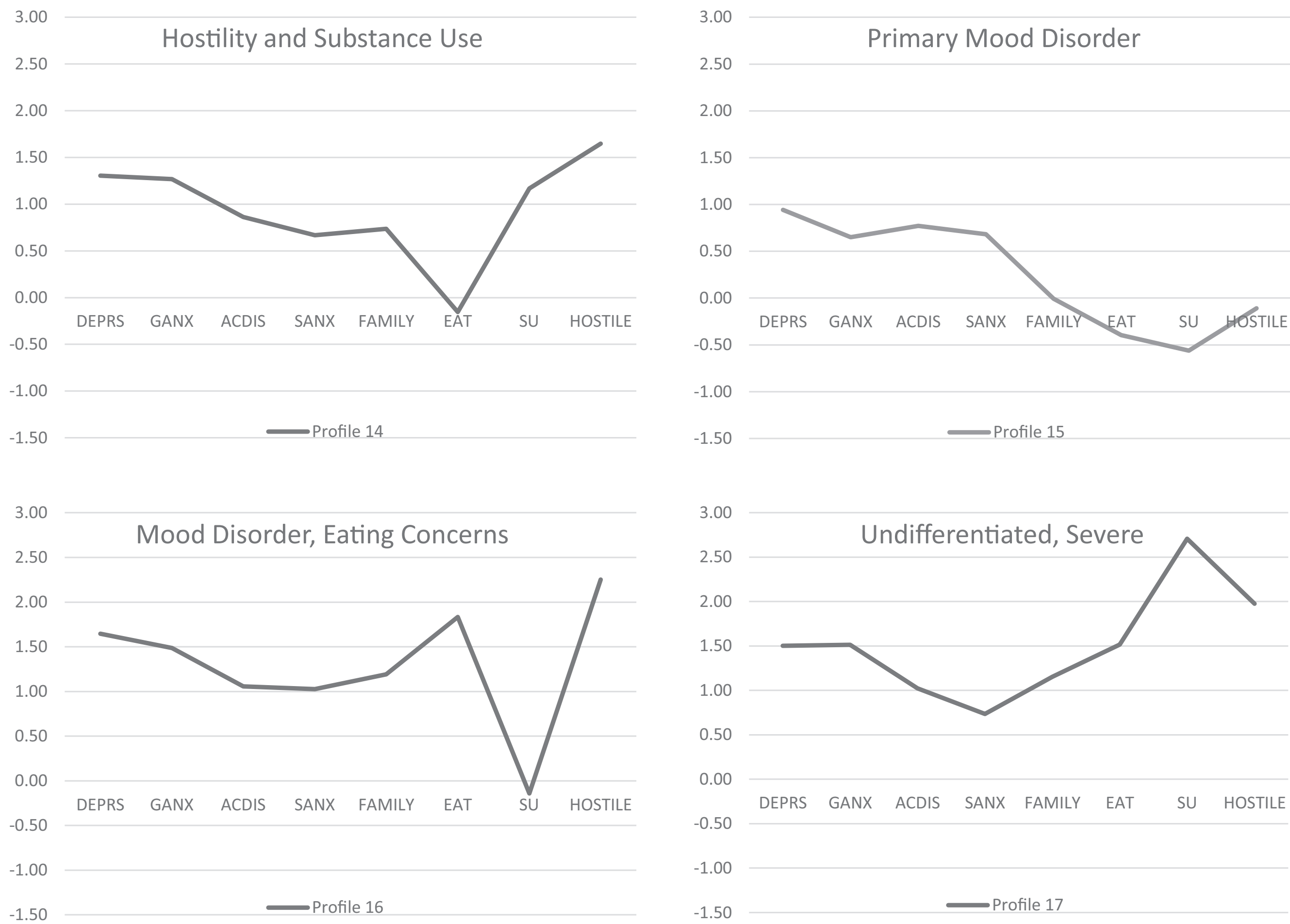

Figure 5. Ungrouped profiles. DEPRS = Depression subscale; GANX = Generalized Anxiety subscale; ACDIS = Academic Distress subscale; SANX = Social Anxiety subscale; FAMILY = Family Distress subscale; EAT = Eating Concerns subscale; SU = Substance Use subscale; HOSTILE = Hostility subscale.

creased reported mood symptoms coincided with a reported history of elevated rates of self-harm, suicidal ideation, and suicide attempts. In contrast, despite nearly identical scores on the Eating Concerns subscales, students with low endorsed mood symptoms are near or below the average for all three risk categories. Additionally, elevated mood symptoms appeared to result in significantly reduced diagnoses of eating disorders despite near identical elevations on the Eating Concerns subscale. Clinicians would be advised to carefully screen for the presence of disordered eating, even if mood symptoms appear to be of greatest concern in the moment.

For the Eating Concerns and Substance Use group, similar endorsed substance use and reported concern about the need to reduce drug and alcohol consumption did not result in similar rates of a substance use disorder diagnosis. Again, co-occurring mood symptoms appear to significantly reduce the rate of substance use diagnosis, despite similar indications of problems. Clinicians would be advised to explicitly assess for substance use disorders in all members of this group. Clinicians would lastly be instructed to assess for past unwanted sexual experiences in members of the Eating Concerns and Substance Use group, as reported rates are significantly in excess of the average.

Students in both eating concerns groups are at or below the average for having deliberately harmed others, in spite of significant substance use in the Eating Concerns and Substance Use group. Clinicians should be advised that substance use in the presence of disordered eating does not appear to indicate an increased likelihood of violence, whereas (see below) for the Primary Substance Use profiles, this is very much the case.

Treatment duration and outcome. With regard to attendance, the two profiles with elevated mood symptoms are the most likely of any in the current study to engage in treatment after one session; however, members of Profile 13 are also among the most likely to cancel or no-show appointments once treatment has begun. One key differentiating feature for feedback on change in treatment of eating concerns is the presence of substance use; the main driver behind our decision to group Profiles 6 and 13. CCAPS feedback should bring clinicians' attention to the indication that co-occurring substance use appears to influence the impact of treatment on the Eating Concerns subscale such that students in the Eating Concerns and Substance Use group have small treatment effect sizes (of d = 0.35 and d = 0.40, respectively). In contrast, students in the Primary Eating Concerns group demonstrate moderate effect sizes roughly twice those of profile 6

($d = 0.61$ and $d = 0.64$) in spite of similar treatment durations and similar starting severity on the Eating Concerns subscale.

With regard to the treatment of co-occurring mood disorder, profiles with elevated mood symptoms in both groups have large effect sizes, and appear to recover well in treatment. Similarly, for the treatment of substance use, both profiles in the Eating Concerns and Substance Use group appear to demonstrate moderate effect size changes. Thus, one critical piece of feedback for clinicians is that treatment as usual appears to effectively improve symptoms of disordered eating for students with or without mood symptoms, but only in the absence of co-occurring substance use. In the latter case, clinicians could be advised to consider referring a student out to an expert, or to strongly consider lengthening the duration of treatment and seeking consultation. Given the havoc an eating disorder may wreak, we consider this a pertinent piece of feedback that may urge clinicians to come up with more intensive or alternative treatment approaches.

Primary Substance Use (Profiles 1, 2, 4, and 11; Figure 2)

This set of profiles described 19.42% of the total assigned sample, and was characterized by elevated substance use and varying degrees of elevated mood symptoms (depression, generalized anxiety, social anxiety, and academic distress). The distinguishing feature of these profiles was that the Substance Use subscale was consistently the most elevated. As noted above, we believe it would be appropriate to combine these profiles into a single grouping, characterized by the relative severity of Substance Use and the Mood subscales.

Assessment and treatment planning. This group appears to provide information that can help clinicians determine the focus of a treatment: Do they address risk and contract for safety foremost, or can they move directly into a treatment focused on reducing substance use? Mood symptoms appear to indicate increased risk. For example, Profile 11 ("severe SU, moderate distress") endorsed rates of self-harm (34.5%) and prior suicide attempt (12.7%) far in excess of Profile 2 ("severe SU, low distress," with 11.8% and 4.6%, respectively), even though both profiles have nearly identical elevations in substance use and endorse a similar amount of binge drinking. It would be important to differentiate this risk of self-harm from risk of violence toward others, which appears driven by elevations on the Substance Use subscale—both profiles indicate a similar elevated risk for deliberately acting violently toward others—each at four times the average, while Profiles 1 and 4 (with far lower substance use) report average rates of violence toward others.

Analyses of diagnosis data reveal that elevated mood symptoms (as seen in Profiles 4 and 11) are markers of students who may be struggling with multiple problems—depression and anxiety alongside substance use. In contrast, members of Profiles 1 and 2, with low mood symptoms, are well below the base rate for every diagnosis other than substance use. Clinically, these findings indicate that for particular groups of clients it is indicated to assess for risk and to work to develop a sound understanding of the interaction between distress symptoms and substance use. For students with higher substance use, it would be important to assess for the potential for violence toward others.

Treatment duration and outcome. With regard to response to treatment, the primary feedback to provide is that all four profiles indicated strong response over relatively brief treatment durations, the length of which appeared driven more by distress symptoms than substance use. Profiles 1 and 2 (with lower mood symptoms) required seven sessions, on average, while those with higher mood symptoms (Profiles 4 and 11) required nine sessions. All four profiles indicated significant change on the Substance Use subscale, with large effect sizes for the most elevated, and moderate sizes for the more moderate elevations. Additionally, those profiles with higher mood symptoms indicated moderate and large effect-size changes on the Depression subscale, respectively. Clinicians could be reassured that, on average, students with primary substance use problems do well in treatment with fewer than 10 sessions, despite their significant Substance Use subscale scores. Important to note is that roughly 25% of members of Profile 1 discontinued treatment after one session; however, once engaged, they were the least likely to cancel or no-show.

Family Concerns and Hostility (Profiles 5 and 8; Figure 3)

Members of this group were distinguished by their report of family concerns and significantly elevated hostility. Profile 5 indicated average distress symptom, while Profile 8 captured elevated mood symptoms. Together, these profiles represent 8.23% of the assigned sample.

Assessment and treatment planning. Both profiles appear to capture students who are feeling unsupported and angry. Students endorsing these patterns also endorsed below-average perceived family support and high levels of having experienced harassing or controlling behaviors from others. The two profiles are significantly above the base rate in both considering harm to others (28.1% and 23.4%, respectively, vs. the base rate of 7.0%) and in having actually harmed another (7.6% and 4.7%, vs. the base rate of 2.0%). The presence of co-occurring mood symptoms appears to reduce the amount of reported harm, considered or perpetuated. Clinicians should be prepared to assess for current thoughts of harming others, as well as for impulsivity and emotion regulation skills.

The presence of mood symptoms in Profile 8 approximately double the likelihood of having self-harmed, or considered or attempted suicide, while Profile 5 has incidences of these risk items at roughly the base rate. Thus, while mood symptoms appear to attenuate somewhat the risk of hostility, they appear also to increase significantly the risk of harm to self. Clinicians would be advised to be on the lookout for these risks, perhaps the result of anger turned toward the self.

Treatment duration and outcome. Both profiles in this grouping appear to have roughly similar treatment durations, despite substantial differences in reported mood symptoms. Between 9 and 10 sessions appears sufficient to effect large changes in reported hostility for Profiles 5 and 8 ($d = -1.20$ and -1.09, respectively). In addition, Profile 8 demonstrated significant improvement in depression symptoms ($d = -1.18$). This could be very helpful information for clinicians who, facing an angry and possibly isolated student, might have good cause for concern. Treatment appears to be quite effective for these students. Regarding the Family Concerns subscale, we were unable to determine the effectiveness of treatment. The CCAPS-34 does not include the Family Concerns subscale and, thus, our repeated-measures methodology lacked sufficient information to calculate an effect size. Students in Profile 8 were highly likely to attend more than one session. Both profiles indicated above average rates of cancellation and no-show, with members of Profile 5 missing nearly one appointment for every two attended. An effective alliance may be difficult to form with members of this profile due to anger or, perhaps, unaddressed concerns related to minority status (see below).

Undifferentiated (Profiles 7 and 10; Figure 4)

Assessment and treatment planning. Profiles 7 and 10 had relatively flat presentations, with no elevations on any particular subscales. This group captured slightly more than 38% of the total sample, and appears to indicate students with adjustment problems and mild distress. This group is substantially below average for all risk items. Profile 10, in particular, appears to represent students with low or no apparent risk. Both profiles may best describe those students long considered the traditional province of college counseling. Diagnosis results corroborate this, with few students garnering a depression diagnosis (12.2% and 3.5%, respectively) or anxiety diagnosis (34.10% and 19.0%, respectively). These students are likely excellent candidates for a short-term, problem-focused model, and may also be excellent referrals to support groups or skills-based groups.

Treatment duration and outcome. The Undifferentiated profiles change little during the course of treatment, likely due to being closer to the floor on the CCAPS subscales than other profiles. Average treatment periods were among the briefest in the sample, with Profile 10 requiring roughly seven sessions. Not surprisingly, students in Profile 7 remained in treatment slightly longer (roughly nine sessions), reflecting their more elevated (if still mild) symptoms. It is difficult to gauge the success or failure of treatment, since scores were so close to the floor. However, we note very little worsening in either profile, based on the clinical cut scores reported in Table 4. Not surprisingly, both profiles had relatively high rates of missed opportunities (cancellations and no-shows). Profile 10 had the highest rate of discontinuation of therapy after one session, likely because low distress provided little motivation to attend treatment.

Ungrouped Profiles (Profiles 14, 15, 16, and 17; Figure 5)

Four profiles were too distinct to reasonably group with others. These we determined to report individually as stand-alone groupings with meaningful differences. Profile 15 was quite common among respondents, and reflected 10.2% of the total assigned sample. The other three profiles were relatively uncommon; all together representing less than 5% of the total assigned sample. All four profiles had elevated mood symptoms, and differed in the associated elevations on the Eating, Substance Use, and Hostility scales.

Assessment and treatment planning. Profile 15 appears to reflect low-risk mood disordered students. This profile is differentiated from the others by average rates of self-harm and prior suicide attempts, in contrast to rates at least twice the base rate in Profiles 14 and 16, and nearly three times the rate in Profile 17. This is particularly notable given that the differences between profiles on the Depression subscale are relatively small—between 0.3 and 0.6 *SD*. Indeed, at this similarly elevated level of mood symptoms, it is the associated elevations that appear to escalate risk. Profile 15 has none, while the other three are elevated on at least one of three subscales: Eating Concerns, Substance Use, or Hostility. Feedback to clinicians would draw attention to these important interactions, indicating that they appear to exacerbate risks usually associated with depressed mood more than would be indicated by the elevation of the Depression subscale alone.

Clinicians should be informed that students in Profile 14 evidence indications of problems with substance use, while those in Profile 16 appear to have difficulties with disordered eating. These are supported by diagnostic analyses which indicate that, while depression is by far the most likely diagnosis for these profiles, alcohol use and eating disorders diagnoses were also given to members of Profiles 14 and 16, respectively. For Profiles 14, 15, and 16, there also appears to be a significant anxiety component, with diagnosis rates close to those for depressive disorders. Importantly, clinicians' attention should be brought to bear on assessing Profile 14 for potential prodromal or psychotic symptoms, as the rate of schizophrenia diagnosis is 12 times the base rate.

We call particular attention to Profile 17, which appears to represent students who are in crisis and intensely distressed. While representing only 1.34% of the total sample, this profile has significant clinical relevance. Notably, according to diagnostic data, 71% of students exhibiting this profile completed their CCAPS-62 as part of a crisis session, rather than a routinely scheduled intake. The majority (roughly 60%) reported a history of self-harm and suicidal ideation, and 22% reported a past suicide attempt—the largest proportion of any profile. Members of this profile were six times more likely to report having intentionally harmed another person (13.7% vs. a base rate of 2.0%). In terms of diagnostics, 25% of students fitting this profile were given a personality disorder diagnosis, and 6.5% a schizophrenia diagnosis—the most of either diagnosis for any profile. Clinicians, alerted to these properties, could attend closely to risk and the possibility of psychosis or characterological challenges. Given the low base rate for this profile, we would also suggest that students exhibiting this symptom pattern be considered for immediate case consultation, rounds, or peer supervision.

Treatment duration and outcome. Our results indicate that there are not substantial differences in change over time for Profiles 14, 15, and 16. All three profiles appear more likely to engage in treatment than the average student, to have roughly similar treatment durations (10 or 11 sessions, on average), and large treatment effect sizes for the Depression subscale despite some differences in the ratio of missed opportunities to attend sessions. This dovetails with a recent benchmarking study (Minami et al., 2009), which found support for effect sizes in college counseling similar to those found in randomized control trials. Feedback on these profiles would indicate that these students tend to do quite well in treatment as usual. Centers with session limits less than 10 would be advised to consider relaxing their restriction. With regard to Profile 14, and the possibility of a schizophrenia diagnosis, the base rate of the diagnosis is so low (6.0%) relative to the sample, that any treatment effects would be lost in our analyses.

Profile 17 required, on average, 13 sessions in treatment—roughly the equivalent of a semester, and nearly three times the national average for counseling centers. Effect sizes for recovery are quite large for this profile, but need to be interpreted carefully. Symptoms of depression seem to meaningfully change ($d = -1.32$) and students appear to end treatment in a similar state as those in other profiles with high levels of depression—that is to say that it appears that treatment is quite effective. On the other hand, symptoms of hostility remain quite high, despite a significant change ($d = -1.21$); more than 86% of students remained above the clinical cut score for hostility posttreatment. Because students are so elevated in their distress, many remain in the distressed range even after months of treatment. Members of Profile 17 also were among the most likely not to pursue treatment after the first session and had the highest ratio of missed opportunities to attended sessions. It is possible that some of the failure to engage is due to hospitalization or immediate referral to a specialist; however, the missed opportunities rate appears to indicate a group of students who are very challenging to engage in treatment.

Clinical guidance for this profile should emphasize the importance of getting these students into treatment quickly, and sustaining that treatment for a duration of more than 3 months. Indeed, the evidence appears to indicate that treatment is effective in reducing distress, but that a substantial amount of treatment may be necessary to affect recovery. Given the high incidence of personality disorder diagnosis, this is unsurprising. Empirically validated treatments for borderline personality disorder, for example, indicate treatment durations of at least 1 year (e.g., Linehan et al., 2006). Extra effort should be taken to engage these students in treatment beyond the first session.

Profiles and Ethnic Minorities

One surprising trend that emerged was that some profiles evidenced a significant overrepresentation of minority students. This has the potential to bring clinicians' attention to students who may be challenged by problems more unique to minority status, and to approach these with particular sensitivity. For example, both Family Concerns and Hostility profiles had disproportionate representation of minority groups. Profile 5 had roughly twice the number of African American and Hispanic students relative to base rates. Profile 8, which was similar to Profile 5 but with elevated mood symptoms, had a different mix of minority representation. Here, Asian American students were represented at twice the base rate, and African American and Hispanic students at roughly half again their respective base rates. Both profiles had above-average ratios of missed opportunities, particularly Profile 5, indicating some difficulty in engaging with treatment. Although speculative, this may reflect the impact of a fragile or weak working alliance, perhaps due in part to some therapists' less-than-optimal responsiveness or attunement to specific cultural needs.

Lastly, Profile 16 included similar overrepresentations to those in Profile 8. All three profiles with overrepresentation of ethnic minorities lacked elevations on the Substance Use subscale. In contrast, some profiles were dominated by Caucasian students. Profiles 1 and 2 had half the African Americans, and a third of the Asian Americans as the base rate. Interestingly, Hispanic students were not underrepresented in these profiles. We would note that elevated hostility is not exclusive to profiles with larger portions of ethnic minorities. Profile 14 (Mood Disorder and Substance Use), for example, has similar significant reported hostility, but reflects an average ratio of minority to Caucasian students.

Two important implications for clinical feedback derive from these findings. Foremost, clinicians should be reminded that minority students can face stressors, both distal and proximal, which can exacerbate vulnerability to particular psychopathology (e.g., Dohrenwend, 2000). Clinicians should be made aware that these profiles are more common among particular ethnic groups, and may reflect important phenomena outside the clinician's own experience. Following from this, clinicians should be encouraged to challenge their assumptions about the nature and source of clinical distress and, potentially, to seek additional peer support in developing a better understanding of the student's culture.

Conclusion

This work is preliminary, but indicates that data-driven methods applied to large sets of clinical information can provide clinically useful feedback for routine care. Multidimensional profiles of students seeking counseling services appear to be statistically reliable and clinically rich. While the current study extracted 16 statistically reliable profiles, our clinical examination indicated that many could be reduced to sets of profiles described by a principle area of distress and meaningful interactions between two or more subscales. This resulted in five groups of profiles, and four ungrouped profiles. Feedback for each set differed based on the type of interaction. This feedback should be possible to incorporate into a set of algorithms for use in CCAPS feedback.

Latent profile analysis, when combined with a rational approach to validation, appears to be a useful means for taking very large datasets and reducing them to subgroups with meaningful clinical implications. The size of our datasets was an advantage and limitation in this study. The limitation, as noted throughout this article, is that statistical significance was practically meaningless as an indicator of validity. The advantage of our very large sets of data was that we had the power to extract several small but clinically relevant profiles that might have been missed in studies with smaller samples. While nearly 50% of the sample was accounted for by three profiles, the remaining 50% was divided into groups no larger than 8% and as small as 1.34%. While small, these groups had robust clinical differences and very important meaning for interpreting CCAPS scores.

Limitations

Our samples were collected several years apart, and one was drawn from the fall semester alone, while the others covered an entire academic year. While it seems fair to conclude that these data were drawn from the same population—college students in counseling—some fluctuations between these datasets could have caused results to cross-validate more poorly than we might expect from samples drawn from the same time period. With regard to replication, we anticipated that many profiles would not match up perfectly since, even in large samples, the proportion of certain types of students could fluctuate. For example, if there were a portion of freshmen with severe and debilitating psychological problems that were worsened by the transition to college, we might expect these students to seek help in the fall semester, and to be largely absent in spring. Since our exploratory sample was collected during only one semester, a group such as this could be overrepresented relative to our confirmatory sample, which represents a full academic year. Fluctuations in a particular response profile could determine whether and how that profile emerges in an LPA. If the portion of the sample fluctuated around 1%, our cut-off would eliminate from our results any profile below 1%, and, thus, change the makeup of the other profiles.

A substantial minority (approximately 30%) of clients analyzed were not given an acceptable classification. Methods such as the ones used in the current study will never fully capture every individual, as there will always be fuzzy boundaries between groups, and clients who fall on those boundaries. With increasingly large datasets, however, it may be possible to accurately extract profiles of clients representing less than 1% of the total. After all, there are disorders with base rates of 1% or lower. It is also important to mention that the determination of diagnosis in the current study lacked tests of reliability and validity and, thus, is a better reflection of what clinicians are seeing and coding than, perhaps, of true and reliable diagnosis. Furthermore, our change data were collected from several settings, with different protocols in place to determine when a CCAPS-34 was administered. As such, it is impossible to be certain that all "post" data points took place on the last session.

Future Directions

Despite these limitations, the results of this study offer an initial indication of the potential for a system of classification and analysis based on patient profiles. By modeling heterogeneity within patient populations, researchers can move closer to validly capturing important characteristics of a new treatment seeker. Such data-driven work has the promise of supporting existing theory or fostering new hypotheses about diagnosis, the interaction of symptoms, and treatment outcome. Translated into clinical tools, this kind of work may offer much-needed empirical support for clinicians and administrators. By moving away from assumptions of homogeneity in naturalistic samples, it appears possible to move closer to meaningfully describing an individual treatment seeker and, through the use of empirically based tools, to augment clinicians' anecdotal experience with empirical findings from tens of thousands of students.

These tools will need substantial testing and adjusting in order to learn how best to provide complex feedback in a clinical setting. The CCMH practice research network offers one venue for testing these tools by applying them to some volunteer test centers and examining the effectiveness relative to other centers (through outcomes data and qualitative feedback). As we have noted, others are attempting such work using different measures and methods, and future research efforts should be made to look for patterns that emerge regardless of method, which may unite our understanding of reporting styles, symptom profiles, and individuals' perceptions of their own problems. The future of this work is not in how results are different, but how they are alike.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Andreasen, N. C., & Grove, W. M. (1982). The classification of depression: Traditional versus mathematical approaches. *American Journal of Psychiatry, 139,* 45–52. http://dx.doi.org/10.1176/ajp.139.1.45

Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77,* 203–211. http://dx.doi.org/10.1037/a0015235

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8,* 338–363. http://dx.doi.org/10.1037/1082-989X.8.3.338

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9,* 3–29. http://dx.doi.org/10.1037/1082-989X.9.1.3

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25,* 6–19. http://dx.doi.org/10.1080/10503307.2013.817696

Cain, N. M., Ansell, E. B., Wright, A. G. C., Hopwood, C. J., Thomas, K. M., Pinto, A., . . . Grilo, C. M. (2012). Interpersonal pathoplasticity in the course of major depression. *Journal of Consulting and Clinical Psychology, 80,* 78–86. http://dx.doi.org/10.1037/a0026433

Castonguay, L. G., Locke, B. D., & Hayes, J. A. (2011). The Center for Collegiate Mental Health: An example of a practice–research network in university counseling centers. *Journal of College Student Psychotherapy, 25,* 105–119. http://dx.doi.org/10.1080/87568225.2011.556929

Center for Collegiate Mental Health. (2010). *CCAPS 2010 technical manual*. University Park, PA: Author.

Center for Collegiate Mental Health. (2015). *2014 annual report* (Publication No. STA 15–30). University Park, PA: Author.

Cohen, J. (1988). *Statistical power analysis in the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

Crow, S. J., Swanson, S. A., Peterson, C. B., Crosby, R. D., Wonderlich, S. A., & Mitchell, J. E. (2012). Latent class analysis of eating disorders: Relationship to mortality. *Journal of Abnormal Psychology, 121,* 225–231. http://dx.doi.org/10.1037/a0024455

Dohrenwend, B. P. (2000). The role of adversity and stress in psychopathology: Some evidence and its implications for theory and research. *Journal of Health and Social Behavior, 41,* 1–19. http://dx.doi.org/10.2307/2676357

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law, 2,* 293–323. http://dx.doi.org/10.1037/1076-8971.2.2.293

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12,* 19–30. http://dx.doi.org/10.1037/1040-3590.12.1.19

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy. Efficacy, effectiveness, and patient progress. *American Psychologist, 51,* 1059–1064. http://dx.doi.org/10.1037/0003-066X.51.10.1059

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167,* 748–751. http://dx.doi.org/10.1176/appi.ajp.2010.09091379

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12.

Kraus, D. R., Castonguay, L., Boswell, J. F., Nordberg, S. S., & Hayes, J. A. (2011). Therapist effectiveness: Implications for accountability and patient care. *Psychotherapy Research, 21,* 267–276.

Leach, C., & Lutz, W. (2010). Constructing and disseminating outcome data at the service level: Case tracking and benchmarking. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence* (pp. 257–283). West Sussex, UK: Wiley. http://dx.doi.org/10.1002/9780470687994.ch10

Linehan, M. M., Comtois, K. A., Murray, A. M., Brown, M. Z., Gallop, R. J., Heard, H. L., . . . Lindenboim, N. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry, 63,* 757–766. http://dx.doi.org/10.1001/archpsyc.63.7.757

Locke, B. D., Buzolitz, J. S., Lei, P. W., Boswell, J. F., McAleavey, A. A., Sevig, T. D., . . . Hayes, J. A. (2011). Development of the Counseling Center Assessment of Psychological Symptoms—62 (CCAPS–62). *Journal of Counseling Psychology, 58,* 97–109. http://dx.doi.org/10.1037/a0021282

Locke, B. D., McAleavey, A. A., Zhao, Y., Lei, P., Hayes, J. A., Castonguay, L. G., . . . Lin, Y. (2012). Development and initial validation of the Counseling Center Assessment of Psychological Symptoms—34. *Measurement and Evaluation in Counseling and Development, 45,* 151–169. http://dx.doi.org/10.1177/0748175611432642

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., . . . Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology, 73,* 904–913. http://dx.doi.org/10.1037/0022-006X.73.5.904

Lutz, W., Martinovich, Z., Howard, K. I., & Leon, S. C. (2002). Outcomes management, expected treatment response, and severity-adjusted provider profiling in outpatient psychotherapy. *Journal of Clinical Psychology, 58,* 1291–1304. http://dx.doi.org/10.1002/jclp.10070

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders, 118,* 60–68. http://dx.doi.org/10.1016/j.jad.2009.01.019

McAleavey, A. A., Nordberg, S. S., Hayes, J. A., Castonguay, L. G., Locke, B. D., & Lockard, A. J. (2012). Clinical validity of the Counseling Center Assessment of Psychological Symptoms—62 (CCAPS–62): Further evaluation and clinical applications. *Journal of Counseling Psychology, 59,* 575–590. http://dx.doi.org/10.1037/a0029855

McAleavey, A. A., Nordberg, S. S., Kraus, D., & Castonguay, L. G. (2012). Errors in treatment outcome monitoring: Implications for real-world psychotherapy. *Canadian Psychology, 53,* 105–114. http://dx.doi.org/10.1037/a0027833

McLachlan, G. M., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley. http://dx.doi.org/10.1002/0471721182

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press. http://dx.doi.org/10.1037/11281-000

Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality, 60,* 117–174. http://dx.doi.org/10.1111/j.1467-6494.1992.tb00269.x

Minami, T., Davies, D. R., Tierney, S. C., Bettmann, J. E., McAward, S. M., Averill, L. A., . . . Wampold, B. E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology, 56,* 309–320. http://dx.doi.org/10.1037/a0015398

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.

Nordberg, S. S., Castonguay, L. G., Fisher, A. J., Boswell, J. F., & Kraus, D. (2014). Validating the rapid responder construct within a practice research network. *Journal of Clinical Psychology, 70,* 886–903. http://dx.doi.org/10.1002/jclp.22077

Nordberg, S. S., McAleavey, A. A., Duszack, E., Locke, B. D., Hayes, J. A., & Castonguay, L. G. (2016). *Better fit, same meaning: Hierarchical models of general distress on the CCAPS-34*. Manuscript submitted for publication.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14,* 535–569.

Owen, J. J., Adelson, J., Budge, S., Kopta, S. M., & Reese, R. J. (2014). Good-enough level and dose-effect models: Variation among outcomes and therapists. *Psychotherapy Research*. Advance online publication.

Stulz, N., & Lutz, W. (2007). Multidimensional patterns of change in outpatient psychotherapy: The phase model revisited. *Journal of Clinical Psychology, 63,* 817–833. http://dx.doi.org/10.1002/jclp.20397

Stulz, N., Lutz, W., Kopta, S. M., Minami, T., & Saunders, S. M. (2013). Dose-effect relationship in routine outpatient psychotherapy: Does treatment duration matter? *Journal of Counseling Psychology, 60,* 593–600. http://dx.doi.org/10.1037/a0033589

Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology, 75,* 864–874. http://dx.doi.org/10.1037/0022-006X.75.6.864

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

Wolf, E. J., Lunney, C. A., Miller, M. W., Resick, P. A., Friedman, M. J., & Schnurr, P. P. (2012). The dissociative subtype of PTSD: A replication and extension. *Depression and Anxiety, 29,* 679–688. http://dx.doi.org/10.1002/da.21946

Wolf, E. J., Miller, M. W., Harrington, K. M., & Reardon, A. (2012). Personality-based latent classes of posttraumatic psychopathology: Personality disorders and the internalizing/externalizing model. *Journal of Abnormal Psychology, 121,* 256–262. http://dx.doi.org/10.1037/a0023237

Youn, S., McAleavey, A. A., Nordberg, S., Castonguay, L. G., Locke, B. D., & Hayes, J. A. (June, 2012). Sensitivity to change of the CCAPS. Paper presented at the annual meeting of the Society for Psychotherapy Research, Virginia Beach, VA.

Received December 29, 2014
Revision received January 22, 2016
Accepted January 27, 2016 ■