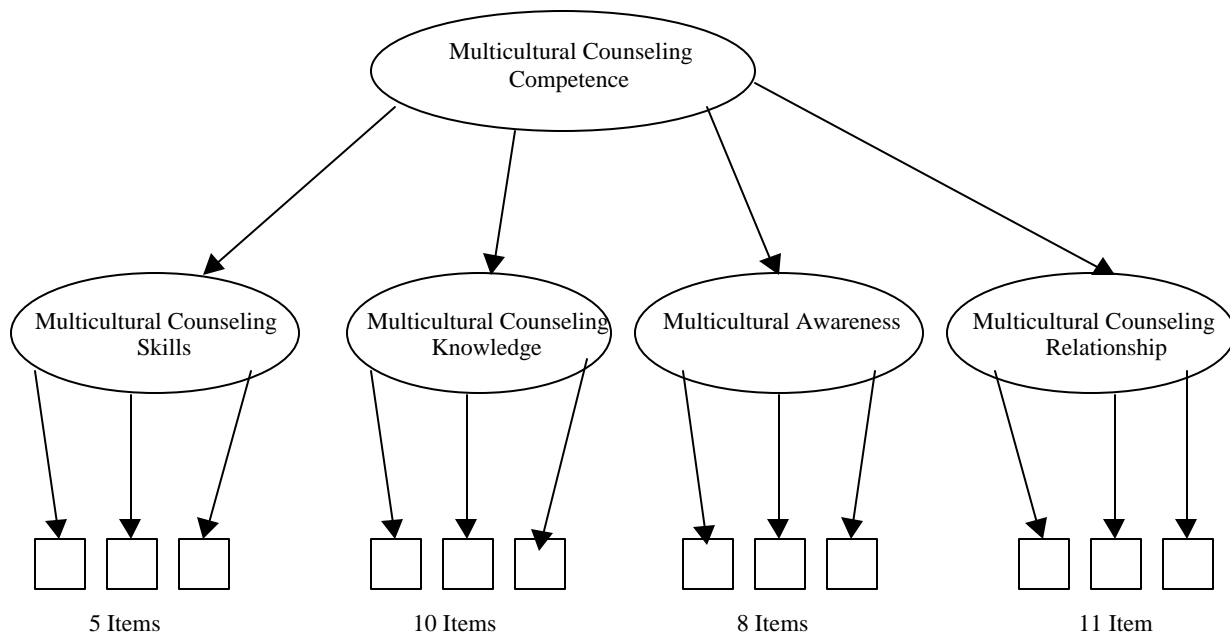


## Assignment: Statistical model for your example The Multicultural Counseling Inventory

The Multicultural Counseling Inventory is a self-report instrument that measures the competence of any counselor working with a client who is a racial, ethnic, or cultural minority. It consists of four subscales: Multicultural Counseling Skills, Multicultural Counseling Knowledge, Multicultural Awareness, and Multicultural Counseling Relationship. It is a 4-point Likert scale: 4 = very accurate, 3 = somewhat accurate, 2 = somewhat inaccurate, and 1 = very inaccurate. Items were worded so that a score of 4 indicated high multicultural competence and a score of 1 indicated low multicultural competence. The items, including the attitudinal and sensitivity items, were behaviorally stated, (e.g., statements began with expressions such as “I am able to,” “I use,” “I am skilled at,” “I am effective with,” “I am comfortable,” “I make,” “I recognize,” and “I am successful at”), and randomly arranged throughout the scale.



Example of items for Multicultural Counseling Skills: “I am able to quickly recognize and recover from cultural mistakes or misunderstanding” and “I use several methods of assessments such as free response questions, observations, and varied sources of information and excluding standardized tests.”

Example of items for Multicultural Awareness: “I am involved in a advocacy efforts against institutional barriers in mental health services for minority clients (e.g., lacking of bilingual staff, multicultural skilled counselors, and outpatient counseling facilities,” and “my life experiences with minority individuals are extensive (e.g., via ethnically integrated neighborhoods, marriage, and friendship).”

Example of items for Multicultural Counseling Relationship: “I find that differences between my worldviews and those of the clients impede the counseling process,” and “I have difficulties communicating with clients who use a perceptual, reasoning, or decision-making style that is different from mine.”

Example of items for Multicultural Counseling Knowledge: “When working with minority clients, I keep in mind research findings about minority clients’ preferences in counseling,” and “I make referrals or seek consultations based on the clients’ minority identity development.”

The Student Model variable is a counselor’s multicultural counseling competency, a total score comprised of four subscales. Each subscale consists of different numbers of items, the observable variables, which purport to measure different dimensions of multicultural counseling

competency. The accumulated scores from the subscales provide the evidence that we will use to make an inference about a counselor's MCT. Likewise, the only evidence that we use to make inferences about the four dimensions of competency is contained in the added scores on items. This assessment is using classical test theory as the base of the statistical model:

$$\text{observed score} = \text{true score} + \text{measurement error},$$

where measurement error is a draw of a random variable with mean zero.

Therefore, the higher the total score on a subscale, the more competent the counselor is on that dimension. For example, a person who scores 18 on the subscale of Multicultural Counseling Skills would be assumed to be more skilled on this dimension than one who scores 10 on that subscale. Likewise, at a more macro level, an accumulated score of 130 from the four subscales indicates a higher competency than would a score of 100.

But what if the "evidence" is not what it appears to be, i.e. what if the credibility, sensitivity, or objectivity of the "evidence" is questionable? If any of these three possibilities should occur, then the "evidence" may very well not support the inference that we would like to make about the latent (i.e., unobservable) variable. For instance, the scale of Multicultural Counseling Skills is designed to measure how skilled a counselor is in working with diverse clients. Instead of observing a counselor's performance in a real counseling session, this assessment uses subjective self-reports to evaluate a counselor's skill level. A high score could be evidence of high competence, but it could also be evidence of something altogether different, because the counselor is not very honest in his or her self-reports or, even if honest, the counselor's perception of his or her multicultural counseling skills is inaccurate. These potential problems of dishonesty and misperception, which would reduce both objectivity and credibility, can be found in each subscale.

Secondly, the issue of social desirability addressed before is also a factor that weakens the relationship between our claims and observed data. Most counselors want to see themselves or at least want to project the image of themselves as multiculturally competent. This is shown by research that has found scores on self-reported instruments to be associated with social desirability (Worthington et al, 2000). Thus, social desirability is an alternative explanation that needs to be discounted before one can make any valid inferences from the results of the self-assessment.

These various potential problems that would limit the accuracy of classical test theory are often interrelated. For instance, social desirability may cause dishonesty that in turn threatens credibility and objectivity. These are important issues to keep in mind both for assessment design and for assessment analysis. For assessment design, we need to be careful in the methods we choose to gather evidence so as to maximize their value in making valid and meaningful inferences about our variables of interest. For assessment analysis, we need to always be alert for and, if we can, rule out alternative explanations that call into question the validity of the inferences we have drawn. Perhaps our "evidence" does not tell us what we thought it did, but paradoxically that does not mean it is useless. Perhaps the data we have gathered can be used as evidence for something we did not predict *a priori*, which could have serendipitous relevance and value for our variables of interest.