
Chapter 3

Convergent and Discriminant Validity:

Treatment Outcome and Process Ratings by Parents & Youth

Forms of Validity

Two forms of validity are required for a valid measure: convergent and discriminant (or divergent) validity according to Campbell and Fiske (1959). Convergent validity requires that measures of similar constructs should be positively correlated. For example, we validated the Symptoms and Functioning Severity Scale (SFSS) score by determining how it correlated with similar measures, namely the Child Behavior Checklist (CBCL; Achenbach, 1991), the Youth Self Report (YSR; Achenbach), the Youth Outcomes Questionnaire (Y-OQ[®]; Wells, Burlingame & Lambert, 1999), and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999). The correlations, approximately $r = 0.80$, suggested that the SFSS was very similar to the other instruments. They are all measures of reported emotional and behavioral problems (see Chapter 4 for more detail).

Discriminant validity is often neglected in describing the validity of measures (Fiske & Campbell, 1992). It requires low correlations for measures of unrelated constructs, or negative correlations for related but opposite constructs. For example if the SFSS showed a high positive correlation with the Therapeutic Alliance Quality Rating (TAQR), this would call its construct validity into question because neither of the theories of alliance and psychopathology posit a strong positive connection between symptoms and alliance (Cronbach & Meehl, 1955). In the case of the Peabody Treatment Progress Battery (PTPB), we are especially concerned that related constructs, such as service satisfaction and therapeutic alliance, are distinct. If we purport that service satisfaction and therapeutic alliance as providing different information, then they should not be highly correlated. We use Pearson correlations to evaluate convergent and discriminant validity.

Convergent Validity

We focused the assessment of convergent validity on the key indicator of treatment progress, the SFSS. To assess the convergent validity (also referred to as concurrent validity) of the SFSS, we included in the psychometric study four established measures used to assess mental health status of youths: the CBCL/YSR; the SDQ; the Y-OQ[®] and the CAFAS. Except for the CAFAS, the appropriate version of each measure was completed by each respondent type with the caregiver version used by clinicians as well as caregivers. The CAFAS, by its design, was completed about the youth by clinicians only.

Unfortunately each of these external measures instructs respondents to consider a different recall period when answering items. For example, the CBCL/YSR asks about the last six months; the Y-OQ[®] the last week; the SDQ, either the last month or the last six months; and the CAFAS, the last three months. We created multiple versions of the SFSS, each with a recall period that matched the other measures so that validity tests would be based on data with equivalent time frames. However we were also interested in learning more about what recall period is appropriate. Given variability among established measures in the field, it was clear that this question has not been resolved. Therefore, we created another version of the SFSS with a 2-week recall period -- a time frame that we think may be the most suitable for assessing youths' symptoms and functioning concurrently with treatment. This 2-week version was paired with each of the external measures (with their usual recall periods). Each region received only one type of pairing (e.g., the 1-week SFSS and the 1-week Y-OQ[®] or the 2-week SFSS and the 1-week Y-OQ[®].) Thus, we were able to assess the concurrent validity of the SFSS when identical recall periods with external measures were considered as well as the validity of the two-week recall relative to the time frames of the other measures.

Twenty-six of the 28 regions participating in the psychometric study completed data for the concurrent validity test. Two of the 28 regions were targeted for the reliability test (test/retest) and were not asked to provide the external validity measures. The 26 regions were divided into five groups based on the number of clients they served and any familiarity they already had with administering any of the external measures. This helped assure an adequate number of cases in each test group as well as eased the burden of completing new forms. Each group received the same external measure paired either with the 2-week SFSS or the SFSS with the equivalent recall period as the external measure. Roughly half of the booklets (to be administered at the 1st session) sent to each region included one of the pairings; half included the other pairing (in addition to the other measures in the 1st booklet). Prior to shipping materials, the envelopes that contained the 1st set of booklets were interleaved so that every other envelope included the same pairing. This procedure helped assure we would receive a balanced number of the two pairings (versions) from each region. Table 3.1 shows the number of regions in each of the five measurement groups and the number of clients in each group about whom youths, caregivers, and clinicians reported.

Table 3.1 SFSS Validity Test: Participation by Measurement Group

| SFSS | External Measure | # of Regions | # of Youth Reports | # of Caregiver Reports | # of Clinician Reports |
|--------------------------------|--------------------------|--------------|--------------------|------------------------|------------------------|
| Validity Test | | | | | |
| 2 week & 1 week | 1-week Y-OQ [®] | 2 | 56 | 47 | 59 |
| 2 week & 1 month | 1 month SDQ | 5 | 84 | 75 | 85 |
| 2 week & 3 Month | 3-month CAFAS | 5 | 135 | 103 | 147 |
| 2 week & 6 month | 6-month SDQ | 8 | 155 | 124 | 162 |
| 2 week & 6 month | 6-month CBCL/YSR | 6 | 159 | 137 | 163 |
| Total for Validity Test | | 26 | 589 | 486 | 616 |

Because ratings of severity can differ significantly depending on the type of rater (youth, adult caregiver, or clinician), we matched the samples by respondent type. The CAFAS, for example, is completed by the clinician. When calculating the correlation, we matched it, therefore, with the clinician version of the SFSS. If versions for several types of respondents were available, we matched each version of the SFSS with the corresponding version of the validity measure.

As can be seen in Tables 3.2 and 3.3, the correlations of the SFSS are very high with the CBCL, the YSR, the different versions of the SDQ, and the different versions of the Y-OQ[®]. These correlations range from 0.71 to 0.89 for the full version and from 0.68 to 0.92 for the short forms. The correlation coefficients are close to the internal reliability estimates of the SFSS, which may be considered the maximum that can be obtained (see Table 4.8). These results provide impressive evidence for the convergent validity of the SFSS. The fact that the SFSS correlates highly with the much longer CBCL and YSR makes it an attractive alternative to those scales if the main goal is to track clinical outcomes over time.

An indirect comparison can be made with the SDQ, as it correlates with the CBCL compared to the correlation of the CBCL with the SFSS (0.86). The correlation of the SFSS with the CBCL is very similar to those found between the SDQ and CBCL. Studies have found correlations of 0.82 (Klasen, 2000) with German caregivers, 0.87 with an English sample (Goodman, & Scott, 1999), and 0.74 with a Dutch sample (vanWidenfelt, Goedhart, Treffers, & Goodman, 2003).

It is noticeable in Tables 3.2 and 3.3 that the correlations with the CAFAS are comparably low (about 0.40). This is true for the CAFAS 5T score as well as the CAFAS 8T score. We could find no published studies that correlated the SDQ or the Y-OQ[®] with the CAFAS. However, low correlations between CBCL and CAFAS (0.22) and YSR and CAFAS (0.24) have been reported by Rosenblatt and Rosenblatt (2002),

but attributed to differences in reporter, with CBCL being the caregiver, the YSR the youth, and the CAFAS the clinician. Bates (2001) also reports comparatively low but somewhat higher correlations of the CAFAS with the CBCL (0.42 to 0.49). The present analysis found similar results even with the same type of reporter. We conclude that the relatively low correlation of the SFSS with the CAFAS is most likely caused by the CAFAS not measuring the same construct as these other instruments.

Table 3.2 SFSS Convergent Validity Estimates for the SFSS-33 Version

| | | SFSS-33 | | | N |
|-------------------------|-----------|---------|-------|-----------|-----|
| | | Youth | Adult | Clinician | |
| CBCL | Adult | | 0.86* | | 115 |
| YSR | Youth | 0.77* | | | 134 |
| Y-OQ[®] | Youth | 0.83* | | | 55 |
| | Adult | | 0.89* | | 44 |
| | Clinician | | | 0.87* | 58 |
| SDQ | Youth | 0.75* | | | 229 |
| | Adult | | 0.79* | | 192 |
| | Clinician | | | 0.71* | 239 |
| CAFAS | Cafas5T | | | 0.42* | 115 |
| | Cafas8T | | | 0.40* | 115 |

*Significant at $p < 0.05$.

Table 3.3 SFSS Convergent Validity Estimates for the Short Forms A and B

| | | SFSS – Short Forms A and B | | | | | |
|-------------------------|-----------|----------------------------|--------|--------|--------|-----------|--------|
| | | Youth | | Adult | | Clinician | |
| | | Form A | Form B | Form A | Form B | Form A | Form B |
| CBLC | Adult | | | 0.84* | 0.85* | | |
| YSR | Youth | 0.73* | 0.80* | | | | |
| Y-OQ[®] | Youth | 0.84* | 0.80* | | | | |
| | Adult | | | 0.83* | 0.88* | | |
| | Clinician | | | | | 0.84* | 0.92* |
| SDQ | Youth | 0.71* | 0.68* | | | | |
| | Adult | | | 0.79* | 0.77* | | |
| | Clinician | | | | | 0.73* | 0.73* |
| CAFAS | Cafas5T | | | | | 0.34* | 0.42* |
| | Cafas8T | | | | | 0.38* | 0.43* |

*Significant at $p < 0.05$.

Discriminant Validity

Of measures in the PTPB, we consider adult caregiver ratings first, correlating the adult caregiver and clinician ratings of youth symptoms and functioning (SFSS) with four treatment process measures (adult caregiver life satisfaction was not used).

- Satisfaction with Services Scale (SSS)
- Therapeutic Alliance Quality Scale (TAQS)
- Treatment Outcome Expectations Scale (TOES)
- Motivation for Youth's Treatment Scale (MYTS)

Table 3.4 shows inter-rater correlations; Cohen's (1992) standard for small/medium/large correlations are $r > 0.10/0.30/0.50$. As a rule of thumb, small and medium correlations are no bar to discriminant validity, but large positive correlations would be a concern.

Table 3.4 Correlations Among Adult-Rated Process and Outcome Total Scores

| | Adult Caregiver SSS | Adult Caregiver TAQS | Adult Caregiver TOES | Adult Caregiver MYTS |
|------------------------|---------------------|----------------------|----------------------|----------------------|
| SFSS - Adult Caregiver | 0.06 | -0.03 | 0.08 | 0.43* |
| SFSS - Clinician | 0.07 | 0.02 | 0.10* | 0.29* |

*Significant at $p < 0.05$.

Table 3.5 shows little correlation between youth severity and treatment process ratings in adult raters, except for treatment motivation, which is higher when symptom scores are higher, a straightforward and rational relationship. This pattern supports the convergent and discriminate validity of the battery for these measures.

For youth self-ratings, there were three outcome instruments:

- Brief Multidimensional Students' Life Satisfaction Scale-CEPI (BMSLSS-CEPI)
- Symptoms and Functioning Severity Scale (SFSS)
- Children's Hope Scale (CHS; Snyder et al., 1997)

And five instruments measuring treatment process:

- Youth's Counseling Impact Scale (YCIS)
- Service Satisfaction Scale (SSS)
- Therapeutic Alliance Quality Scale (TAQS)
- Treatment Outcome Expectations Scale (TOES)
- Motivation for Youth's Treatment Scale (MYTS)

Table 3.5 Correlations Among Youth-Rated Process and Outcome Total Scores

| Scales | YCIS | SSS | TAQS | TOES | MYTS |
|--------|-------|-------|-------|-------|--------|
| BMSLSS | 0.24* | 0.16* | 0.23* | 0.17* | -0.10* |
| SFSS | -0.03 | 0.05 | -0.07 | 0.04 | 0.42* |
| CHS | 0.24* | 0.12* | 0.18* | 0.09* | 0.20* |

*Significant at $p < 0.05$.

As presented in Table 3.5, life satisfaction and hope ratings had small correlations with process scores. Symptom scores, on the other hand, were not correlated with process measures except for the expected positive correlation with treatment motivation (MYTS). This correlation was expected because higher treatment motivation usually accompanies serious symptoms.

Inter-Rater Issues

SFSS Inter-Rater Correlations

The inter-rater correlations on the SFSS Total Score had medium-sized effects as shown in Table 3.6.

Table 3.6 Inter-Rater Correlations for SFSS Total Scores

| | SFSS-Adult Caregiver | SFSS-Clinician | SFSS-Youth |
|-----------------|----------------------|----------------|------------|
| Adult Caregiver | 1 | | |
| Clinician | 0.44* | 1 | |
| Youth | 0.45* | 0.36* | 1 |

Notes: For Pearson r , Cohen defines small/medium/large as 0.10, 0.30, 0.50.

*Significant at $p < 0.05$.

While the youth-clinician correlation appears lower than the other two, this difference was not significant ($prob_{CY} = 0.12$, $prob_{AY} = 0.08$). It is typical to obtain correlations about $r = 0.20$ to 0.30 between caregiver and youth (De Los Reyes & Kazdin, 2005). The clinician-adult caregiver correlation (0.44) and the adult caregiver-youth correlation (0.45) are significantly higher than $r = 0.30$ ($p < 0.01$) but the youth-clinician correlation (0.36) is not ($prob = 0.10$).

Test with Three Items (3 Raters)

Instead of regretting the low caregiver-youth correlations, we viewed the SFSS as a test with three items, youth, adult caregiver, and clinician. This “meta-test” had a moderate reliability of $\alpha = 0.68$. The item-total correlations (see Table 3.7) would be good for the usual items in a test. While this viewpoint doesn’t solve the problem of low inter-

rater correlations, it does contradict the idea that adult caregiver and youth ratings are unrelated.

Table 3.7 Item-Total Correlations for SFSS Informants

| Item-Total r | Item |
|----------------|--|
| 0.49 | Adult Caregiver SFSS Externalizing 14-item |
| 0.42 | Clinician SFSS Externalizing 14-item |
| 0.49 | Youth SFSS Externalizing 14-item |

Adult versus Youth Discriminant Validity

We hypothesized that discriminant validity would be somewhat better for adults, because of the less differentiated cognitive and emotional development of the youth. We tested this with matching youth-adult correlation matrices including only the instruments that both respondents shared. According to the results presented in Table 3.8, youth appear to have a more global, less differentiated, view of service satisfaction, alliance, and treatment expectations.

Table 3.8 Battery Instrument Correlations for Youth and Adult Caregivers

| | Scale | SSS | TAQS | TOES | MYTS | SFSS |
|--------------------------------------|-------|--------------|--------------|--------|-------|------|
| Youth | SSS | 1 | | | | |
| | TAQS | 0.70* | 1 | | | |
| | TOES | 0.54* | 0.57* | 1 | | |
| | MYTS | 0.41* | 0.35* | 0.44* | 1 | |
| | SFSS | 0.05 | -0.07 | 0.04 | 0.42* | 1 |
| Adult Caregiver | SSS | 1 | | | | |
| | TAQS | 0.61* | 1 | | | |
| | TOES | 0.32* | 0.31* | 1 | | |
| | MYTS | 0.33* | 0.30* | 0.35* | 1 | |
| | SFSS | 0.06* | -0.03 | 0.08 | 0.43* | 1 |
| Youth and Adult Caregiver Difference | SSS | 1 | | | | |
| | TAQS | -0.15* | 1 | | | |
| | TOES | -0.27* | -0.33* | 1 | | |
| | MYTS | -0.09* | -0.05 | -0.10* | 1 | |
| | SFSS | 0.00 | -0.04 | 0.04 | 0.01 | 1 |
| Significance of Difference | SSS | 1 | | | | |
| | TAQS | 0.03 | 1 | | | |
| | TOES | 0.01 | <0.01 | 1 | | |
| | MYTS | 0.16 | 0.47 | 0.14 | 1 | |
| | SFSS | 0.95 | 0.58 | 0.52 | 0.82 | 1 |

Notes: For Pearson r , Cohen defines small/medium/large as 0.10, 0.30, 0.50.

*Significant at $p < 0.05$.

References

- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.
- Bates, M. P. (2001). The Child and Adolescent Functional Assessment Scale (CAFAS): Review and current status. *Clinical Child and Family Psychology Review*, 4, 63-84.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131(4), 483-509.
- Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin*, 112, 393-395.
- Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(5), 791-799.
- Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist; Is small beautiful? *Journal of Abnormal Child Psychology*, 27, 17-24.
- Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W. et al. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *European Child and Adolescent Psychiatry*, 9, 271-276.
- Rosenblatt, A. & Rosenblatt, J. A. (2002). Assessing the effectiveness of care for youth with severe emotional disturbances: Is there agreement between popular outcome measures? *Journal of Behavioral Health Services & Research*, 29(3), 259-273.
- Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., et al. (1997). The development and validation of the Children's Hope Scale. *Journal of Pediatric Psychology*, 22, 399-421.
- Van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry*, 12, 281-289.
- Wells, M. G., Burlingame, G. M., & Lambert, M. J. (1999). Youth Outcome Questionnaire. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (2nd ed.). Mahwah NJ: Lawrence Erlbaum