

Comparison of computational methods for the evaluation of learning potential in schizophrenia

JOANNA M. FISZDON,^{1,2} AND JASON K. JOHANNESSEN^{1,2}

¹Psychology Service, VA Connecticut Health Care System, West Haven, Connecticut

²Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut

(RECEIVED November 18, 2009; FINAL REVISION March 2, 2010; ACCEPTED March 2, 2010)

Abstract

Learning potential (LP) refers to the ability to improve cognitive performance as a result of training. It is typically assessed by test-train-test administrations of a task, wherein changes in pre-post performance are an index of LP. In schizophrenia research, LP has been suggested as a mediator of the relationship between static neurocognition and functional outcome. While a number of studies do indicate that LP assessment improves prediction of functioning beyond standard administrations of the same task, multiple approaches of computing LP indices have been used in this work. Multiple psychometric issues have been raised with respect to computation of change scores, but have not been widely recognized in LP assessment. To address this issue, the current study aimed to evaluate the test-retest reliability, interrelatedness, construct, and criterion validity of several conventional indices of LP, obtained from a test-train-test version of a list-learning task administered to 43 individuals with chronic schizophrenia. Overall, test-retest and intercorrelation coefficients indicated variable reliability and convergence across methods. While LP indices generally correlated more highly with independent measures of neurocognition and community functioning than pretraining list learning scores, coefficients were comparably small. Recommendations and measurement issues inherent to the LP construct are discussed. (*JINS*, 2010, *16* 613–620.)

Keywords: Psychotic disorders, Neuropsychology, Cognition, Psychometrics, Learning, Adult

INTRODUCTION

The term “learning potential” was first coined in 1964 by Budoff and colleagues (Budoff & Friedman, 1964). Unlike static, single time-point assessments of cognitive ability, learning potential (LP) refers to the ability to improve cognitive performance as a result of training, and entails multiple time-point assessments, with intervening training on the domain of interest. LP assessments were originally developed for use with educable mentally retarded (EMR) children in order to better evaluate which of those with low intelligence quotient (IQs) had an ability to benefit from training or instruction. More recently, LP assessments have also begun to be used in schizophrenia research. Interest in LP in schizophrenia developed as a result of its’ suggested role as a mediator between static neurocognition and functional outcome (Green, Kern, Braff, & Mintz, 2000). Construct validity of LP assessment has been supported by studies indicating that

groups characterized by differences in LP exhibit distinct cognitive profiles (Kurtz & Wexler, 2006; Rempfer, Hamera, Brown, & Bothwell, 2006; Vaskinn, et al., 2008; Wiedl, Schottke, & Calero-Garcia, 2001), as well as distinct cerebral metabolite levels (Ohrmann et al., 2008; Pedersen, Wiedl, & Ohrmann, 2009). A number of studies also indicate that LP predicts functioning, including performance on medication management and problem-solving skill training (Wiedl, 1999; Wiedl et al., 2001), training gains in cognitive rehabilitation training (Wiedl & Wienobst, 1999), work skill acquisition (Sergi, Kern, Mintz, & Green, 2005), readiness for rehabilitation (Fiszdon et al., 2006), performance on a proxy measure of functioning (Kurtz & Wexler, 2006), and the outcome of vocational rehabilitation (Watzke, Brieger, Kuss, Schoettke, & Wiedl, 2008; Watzke, Brieger, & Wiedl, 2009). It should be noted, however, that there are also several reports indicating that LP may not confer an advantage above and beyond the predictive value of static neurocognitive measures (Kurtz, Jeffrey, & Rose, 2010; Tenhula, Kinnaman, & Bellack, 2007; Vaskinn et al., 2008; Woonings, Appelo, Kluiter, Slooff, & van den Bosch, 2003). Furthermore, although many researchers report that LP status is a good predictor

Correspondence and reprint requests to: Joanna Fiszdon, Ph.D., Psychology Service (116B), VA Connecticut Healthcare System, 950 Campbell Ave., West Haven, CT 06516. E-mail: Joanna.fiszdon@yale.edu

of functioning, multiple approaches have been used to determine LP status, and no consensus has yet been reached about the best method for assessing and quantifying LP.

A number of different testing and computational approaches have been used to quantify LP. Studies of LP in schizophrenia have most frequently relied on test-train-test versions of the Wisconsin Card Sorting Test (WCST; Heaton, 1981), a test of executive function, though also used have been list-learning tasks that assess verbal learning and memory. Based on pre-training test performance and relative change in post-training performance, one approach of applying LP assessment involves classifying individuals into learner categories (see Schoettke, Bartram, & Wiedl, 1993 for algorithm), namely, high scorers (those with good pre-training performance), learners (those whose post-training performance is above limits established by reliable change criterion), and nonretainers (those whose post-training performance is within limits established by reliable change criterion). While this approach has the advantage of separately examining distinct LP groups, as with any categorical approach, it reduces statistical power. An alternative to the categorical approach is the continuous gain approach, which relies on a ratio of actual pre-post improvement to maximum possible pre/post improvement (see Sergi et al., 2005 for algorithm). While this approach has the benefit of increasing statistical power, it assumes a linear continuum of learning potential ability. Moreover, because gain scores are scaled relative to optimum performance, individuals with initial performance near optimum can receive very high or low gain scores based on minute differences in pre- and post-training performance – small decrements from initially high performance produce gain scores similar to those of individuals with low pre- and post-training performance, while small improvements in post-training performance produce gain scores similar to those of individuals with poor pre-training performance who make considerable gains following training. An alternative to these two LP indices is the optimized performance approach (see Woonings et al., 2003). As the name suggests, this approach simply relies on the post-training performance, without reference to pre-training performance or amount of improvement as a result of training. While only a “static” post-training score is used, this approach appropriately falls within the realm of learning potential assessment, as the post-training score in effect incorporates the potential effects of the training intervention. While the optimized approach does provide an index of how well an individual can perform following training, it is limited by not directly incorporating pre-training performance, thereby not taking into account the amount of change as a result of training. Two additional indices have also been used in the assessment of LP: a simple difference score between pre-training and post-training performance, and a regression residual score, calculated by regressing pre-training performance on post-training performance (see Weingartz, Wiedl, & Watzke, 2008 for algorithm). While not unique to LP assessment in schizophrenia, the use of varied measurement strategies represents a serious problem for researchers for

several reasons. For one, methodological differences can lead to substantially different conclusions, even using the same data, as well as decrease the potential for the generalization of findings. Methodological differences may also complicate research conclusions and practical implications of LP assessment, as LP computational approaches may differ in basic psychometric (i.e., reliability) properties and inadvertently alter the construct being assessed.

We are aware of only two studies that have examined the psychometric properties of different approaches to LP assessment in schizophrenia. Waldorf and colleagues (Waldorf, Wiedl, & Schottke, 2009) examined the comparability of three different reliable change indices as applied to categorical LP classification using the WCST. While their results indicated particularly high ($\kappa > 0.90$) classification concordance rates for two of the three indices, concordance rates among all three indices were still in the acceptable range ($\kappa > 0.70$). As noted by the authors themselves, while these analyses allowed for a comparison of the three indices, they did not address questions of the validity of the resulting classifications. In a more extensive evaluation of LP indices obtained using the test-train-test administration of the WCST, Weingartz and colleagues (Weingartz et al., 2008) evaluated the test-retest stability, intercorrelations among, and construct validity of several continuous approaches to LP measurement, which included pre-post differences, gain scores, and regression residuals. Also examined was a categorical approach to LP classification, as well as raw pre- and post-training test values. These authors concluded that from among the indices examined, categorical LP classification, raw post-training test performance, and regression residuals had the highest stability and validity. Close examination of these indices, however, reveals low levels of one-year test-retest stability (0.32 to 0.51). Analyses ascertaining concurrent and prognostic validity (assessed with functioning scales and employment status) yielded significant but small (0.25 to 0.35) correlation coefficients for post-training test performance and regression residuals.

Both of these psychometric investigations relied on scores obtained from test-train-test versions of the WCST. An important characteristic of the WCST training is that it focuses on whether, and how quickly, the examinee learns the sorting principles. Once these sorting principles are known, they can be applied time and again to successive administrations of the same task. Particularly in cases where test-retest stability of LP is assessed, it is possible that WCST retest performance may be “spoiled” for individuals who learn the sorting principles during the initial LP training trial – the test may change from one of problem solving ability, to one measuring memory of test rules. In fact, there is data to indicate that the construct validity of WCST performance changes as a result of training. This is demonstrated by changes in the pattern of intercorrelations between pre-training and post-training WCST performance and performance on other cognitive tasks, with a significant increase in the correlation between memory function and WCST performance following training (Wiedl, Schottke, Green, & Nuechterlein, 2004).

An alternative measure that has been used to assess learning potential is word list learning. These tests typically consist of multiple administrations of a word list, with instructions, additional practice, and feedback between the initial and final list recalls. While there is evidence of construct overlap between WCST-based LP and verbal memory assessed by list-learning tasks (Wiedl et al., 2004), unlike with the WCST, there is no rule that, if known by the examinee, will result in perfect list learning performance, and both pre- and post-training performance are an index of the same cognitive domain, verbal learning and memory.

The aim of the current investigation was to examine the reliability and validity of LP indices based on a list-learning task. We evaluated several LP indices based on a test-train-test version of a list-learning task, in which the training phase focused on teaching the use of semantic grouping, an encoding strategy that is associated with greater list recall. Unlike the WCST LP administrations, different versions of the word lists were administered at pre-test, training, and post-test, so the pre-post training differences are more apt to reflect the application of the trained strategy rather than simple practice effects associated with repeated exposure to the same stimuli. Like Weingartz and colleagues (Weingartz et al., 2008), our analyses focused on four psychometric issues: test-retest stability, intercorrelations of different LP indices, construct, and criterion validity.

METHODS

Participants

Participants in the study were 43 individuals with a *DSM-IV* (American Psychiatric Association & Task Force on *DSM-IV*, 1994) diagnosis of schizophrenia ($n = 35$) or schizoaffective disorder ($n = 8$), as assessed by the Structured Clinical Interview for *DSM-IV*, (SCID; First, Spitzer, Gibbon, & Williams, 1996) administered by doctoral-level psychologists. All participants were enrolled in a larger, ongoing, randomized trial of cognitive remediation being conducted at a Veterans Administration Medical Center. For the current analyses, data were included only for individuals who completed the learning potential assessment at both study intake and at two-month follow-up. The study was approved by the local institutional review board (IRB), and all participants signed informed consent.

All participants met the following inclusion criteria: age between 18 and 65 years, no substance abuse in the past month, no known neurological condition that may affect cognition, no documented mental retardation, no medication changes in past month, and no housing changes in past month. The sample was 56% Caucasian and 75% male. Sixty-one percent of the participants had never been married. On average, age at illness onset was 22.08 ($SD = 6.01$), with age at study entry of 49.56 ($SD = 8.47$). Participants had completed an average of 12.53 ($SD = 2.04$) years of education and had a Wechsler Abbreviated Scale of Intelligence (WASI, Wechsler, 1999) IQ estimate of 94.37 ($SD = 14.94$).

They had had an average of 16.27 ($SD = 25.82$) hospitalizations, a Global Assessment of Functioning (GAF; American Psychiatric Association & Task Force on *DSM-IV*, 1994) score of 39.00 ($SD = 6.38$), and a Positive and Negative Syndrome Scale (PANSS; Kay, Fiszbein, & Opler, 1987) score of 54.19 ($SD = 11.43$).

Procedures

All participants in this sample of convenience completed comprehensive intake assessments consisting of psychiatric interviews and cognitive testing, after which they were randomized (1 to 2 ratio) to receive two months of either treatment as usual or cognitive remediation. Cognitive remediation consisted of up to 40 hours of both computerized and paper-and-pencil training, akin to that used in Neurocognitive Enhancement Therapy (NET; see Bell, Bryson, Greig, Corcoran, & Wexler, 2001) and Cognitive Remediation Therapy (CRT; see Delahunty, Reeder, Wykes, Morice, & Newton, 2001), respectively. At the end of the two-month active phase, all participants were again administered psychiatric interviews and cognitive testing.

Measures

Learning potential assessment

A modified test-train-test administration of the California Verbal Learning Test-II (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) was used to assess learning potential. For details of this procedure, the reader is referred to Fiszdon and colleagues (Fiszdon et al., 2006). Briefly, for the standard CVLT-II administration, the examinee is read and asked to recall a list of 16 words belonging to four semantic categories. Each correctly recalled word earns one point. The list is administered a total of five times, for a maximum score of 80. The LP modified CVLT involves the administration of three different word lists (which we refer to as pre-training, training, and post-training lists). While the first administration uses standard procedures, the second list administration is preceded by a training component. Standard administration is again done for the third CVLT list. Immediately prior to and during the administration of the second (training) list, participants were shown how using semantic grouping can lead to better recall, were asked to recall the specific list categories at the end of each recall trial, and were given corrective feedback about list categories as needed (training script available upon request from the first author, J.F.).

The following raw scores and learning potential indices were calculated based on trials 1–5 total scores (maximum of 80) for the pre-training and post-training list administrations: (1) Pre-training score, (2) Post-training score, (3) Categorical LP index (“learner,” “nonlearner,” and “high-scorer” groups) based on pre-training performance and whether post-training performance is outside a 90% confidence interval set around a hypothetical parallel test score

(for algorithm, refer to Schoettke et al., 1993), (4) regression residuals (obtained by regressing pre-training performance on post-training performance; see Weingartz et al., 2008 for algorithm), (5) Post-training minus pre-training difference score, and (6) gain score (ratio of actual pre-post training improvement to maximum possible pre-post training improvement, with maximum possible pre-post improvement being 80 minus pre-training performance; for algorithm see Sergi et al., 2005). The modified CVLT-II was administered at study intake and at 2-month follow-up, and separate LP indices were computed for each time point.

Construct validity measure

The Wisconsin Card Sorting Test (WCST; Heaton, 1981) is a neuropsychological test of abstract problem solving that has previously been adapted to LP assessments, and has been shown to share variance with list-learning performance (Wiedl et al., 2004). The WCST was administered at intake and used in this analysis as a measure of concurrent validity of the learning construct. Standard scores for two of the more common indices, percent errors, and percent perseverative errors were used in the current analyses.

Criterion validity measures

The Quality of Life Scale (QLS; Heinrichs, Hanlon, & Carpenter, Jr., 1984) was administered at intake and at 2-month follow-up and was used to assess concurrent and predictive validity of the LP measures in relation to global functioning. The QLS is a well-validated scale consisting of 21 items organized into four domains: intrapsychic foundations, interpersonal relations, instrumental role function, and common objects and activities. Likert-type (0–6) ratings are made following a clinician-administered interview, with higher total scores indicating better overall functioning. All ratings were made by the first author (J.F.), who was blind to learning potential classification at the time of interview.

Data Analysis

All data were analyzed using the Statistical Package for the Social Sciences (SPSS 17). To determine test-retest stability, bivariate correlations were computed between intake and 2-month LP indices. To determine whether the main study intervention affected the measurement of LP at 2-month follow-up, test-retest analyses for the LP indices were repeated, controlling for the number of cognitive remediation sessions. Next, intercorrelations between the LP indices were computed. Construct validity of LP indices was evaluated using correlations of each LP index with WCST performance. Concurrent and predictive criterion validity was assessed by correlating intake LP indices with intake and 2-month QLS scores. For the categorical LP index variable, all correlations were computed using Kendall's tau-b. For the remaining variables, Pearson correlations were used. Differences in the strength

of correlations between each LP index and the construct and criterion validity measures (WCST, QLS) were tested using William's T2 statistic for dependent correlations (Steiger, 1980). Using William's formula, raw correlation coefficients obtained for each LP index were entered in a pair-wise contrast with the pre-training CVLT score. In this manner we sought to determine whether LP administration of the CVLT provides a stronger measure of association than standard administration, and, if so, which of the LP indices have superior predictive power. These tests were conducted using $\alpha = .05$, two-tailed.

RESULTS

Results for bivariate test-retest correlations were strikingly similar to results for partial correlations that controlled for the number of cognitive remediation sessions (see Table 1). Because we found no indication that the reliability of LP measures was affected by cognitive remediation in this sample, partial correlations are not reported, and all subsequent analyses were conducted without controlling for the number of cognitive remediation sessions. Data on test-retest reliability, intercorrelations among intake LP indices, construct and criterion validity of the LP indices are reported in Table 2.

Test-Retest Reliability

Correlations between intake and 2-month LP indices ranged between $r = .749$ and $r = .105$, with the highest values for the pre-training and post-training scores, significant moderately-sized correlations for categorical LP index, regression residual, and pre-post difference score, and small nonsignificant correlations for gain score.

Intercorrelations Among LP Indices

Post-training scores correlated significantly with all other LP indices with the exception of gain scores. Regression residuals correlated significantly with all other LP indices with the exception of pre-training score. Regression residuals, pre-post difference scores, and gain scores were highly correlated with each other.

Table 1. Intake to 2-month test-retest correlations

	Bivariate correlation	Partial correlation
CVLT List 1 (pre-training)	0.749***	0.713***
CVLT List 3 (post-training)	0.710***	0.721***
Categorical Learner group	0.448***	0.456*
Regression residual	0.342*	0.397**
Post-training–pre-training difference	0.614***	0.585***
Gain score	0.105	0.073

Note. Correlation coefficients of bivariate analyses with statistical significance indicated as follows: * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

Table 2. Test-retest correlations, intercorrelations, and construct validity of learning potential (LP) indices

	Pre-training	Post-training	Categorical LP Index ¹	Regression Residual	Pre-Post Difference	Gain Score
Test-Retest Reliability						
2-Month Follow-up	.749***	.710***	.448**	.342*	.614***	.105
Intercorrelation						
Pre-training	1	.754**	.442**	0	-.266	-.434**
Post-training		1	.563**	.657**	.434**	.225
Categorical LP Index			1	.348*	.243	.161
Regression Residual				1	.964**	.840**
Pre-Post Difference					1	.925**
Gain Score						1
Construct Validity						
WCST % Errors	.385*	.430**	.378**	.214	.104	-.067
WCST % Pers Err	.393**	.438**	.370**	.216	.104	-.078
Concurrent Criterion Validity						
QLS intake	.113	.405**	.251*	.488**	.440**	.375*
Predictive Criterion Validity						
QLS 2-Month	.084	.326*	.205+	.400**	.363*	.359*

Note. Correlation coefficients of bivariate analyses with statistical significance indicated as follows: + $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$. Test-retest reliability for each score is analyzed in CVLT LP administrations conducted at intake and 2-month follow-up. Intercorrelations among LP indices were computed across intake data only. Concurrent validity was assessed in relation to Wisconsin Card Sort Test (WCST) % errors and % perseverative errors and Quality of Life Scale (QLS) ratings at intake. Predictive validity was assessed in correlations between intake LP indices and QLS ratings conducted at 2-month follow up.

¹Kendall's tau b used for correlations with the categorical LP measure, Pearson r used for all other analyses.

Construct Validity

Construct validity was assessed in relation to cognition (WCST % errors and % perseverative errors). Only pre-training score, post-training score, and categorical LP index correlated significantly with WCST performance. A statistical comparison of the strength of these correlations determined that none of the LP indices produced coefficients that differed significantly from pre-training performance. Trend level effects were detected for the gain score method, indicating weaker association with both WCST % errors, $t(40) = -1.79$, $p < .10$, and WCST % perseverative errors, $t(40) = -1.87$, $p < .10$, than pre-training score.

Concurrent Criterion Validity

In contrast to the pattern of results for WCST score, each of the LP indices correlated significantly with intake QLS, despite a small, nonsignificant, correlation with pre-training score. Compared statistically, only one of these LP indices, post-training score, produced a correlation with intake QLS that was significantly higher, $t(40) = 3.04$, $p < .01$, than that of pre-training score. A nonsignificant trend in the same direction was detected for the regression residual method, $t(40) = 1.88$, $p < .10$.

Predictive Criterion Validity

This was assessed in relation to QLS ratings conducted 2 months following intake LP assessment. Across pre-training and LP indices, the pattern of correlations was similar to that

observed with intake QLS, with the exception that the correlation between categorical LP grouping and QLS dropped to trend level. Compared statistically, only post-training score produced a correlation with 2-month QLS that was significantly higher, $t(40) = 2.39$, $p < .05$, than that of pre-training score.

DISCUSSION

The aim of the current study was to evaluate the reliability and validity of several LP indices, as applied to test-train-test administrations of a list-learning measure, the CVLT-II. With regard to test-retest reliability, our results indicate that raw CVLT test scores (pre-training score and post-training score) have high 2-month test-retest reliability (0.70 range). In contrast, test-retest reliability of LP measures based on these scores is variable and modest overall, with coefficients ranging from $r = .34$ to $r = 0.61$ for categorical LP index, regression residual, and pre-post difference score methods. A final method, the gain score, had the lowest test-retest reliability, with a coefficient only marginally exceeding 0. Given that all LP scores were computed based on the same pre- and post-training CVLT scores, which themselves appeared to be highly reliable under repeat administration, this analysis demonstrates how basic psychometric properties of LP scores can be altered by their computational method. In line with the well-known adage about reliability being a necessary but not sufficient condition for validity, our results suggest that gain scores are less than optimal indices of LP. However, we recognize that reliability can vary widely across samples, and our findings require replication in larger

samples before definitive conclusions can be drawn about the usefulness of this particular LP index. Additionally, any conclusions based on these reliability coefficients assume that learning potential is a stable construct that can be assessed reliably over multiple dynamic test administrations, which has been questioned by some (Weingartz et al., 2008). Specifically, it has been suggested that LP may have state-like characteristics, and may be influenced by variables such as psychopathology. If this were the case, one would not expect high test-retest correlations. However, given that from among all of the LP indices studied, only gain scores fail to show significant test-retest correlations, it is more likely that this is in fact due to the psychometric properties of this particular method of LP calculation.

In the absence of a gold standard with which to correlate the various LP measures, it is difficult to draw conclusions about validity based on the pattern of intercorrelations, as they only indicate similarities between classifications, and not necessarily the degree of overlap with the latent construct of interest. It is of interest, however, that post-training performance most consistently correlates with all the other LP indices, suggesting that it best taps the shared variance of the measures and has the highest overlap with the construct being measured by the other indices. Also noteworthy is the high level of agreement among gain scores, pre-post training difference scores, and regression residuals. Given that all three of these measures are based on pre- and post-training performance, this high degree of relationship could be due to a measurement (psychometric) artifact, with high correlations due to shared measurement error across the measures (Linn & Slinde, 1977).

Concerning the construct and criterion validity of LP assessment, several patterns of association between LP indices and independent measures of neurocognitive (WCST) and community (QLS) functioning are notable. First, we found that only post-training performance and the categorical LP index correlated significantly with WCST performance. Because the WCST is a measure of abstract problem-solving ability in which the examinee can improve performance by using corrective feedback, one would expect overlap between WCST scores and learning capacity. Furthermore, correlations between pre-training CVLT score and both WCST scores were significant, suggesting some common variance in the neurocognitive domains assessed by these tests. However, no appreciable relationship with WCST performance was observed for LP indices computed by regression residual, pre-post difference, or gain score methods. For our criterion validity measure, intake QLS, the use of LP indices did appear to add unique and meaningful variance in predicting functioning over the standard (pre-training) CVLT administration, a finding similar to that reported by Weingartz and colleagues (Weingartz et al., 2008). Correlations between intake QLS and post-training score, categorical LP classification, pre-post training difference score, regression residual, and gain score were all statistically significant, while the correlation between QLS and pre-training CVLT score was not. A very similar pattern of results in relation to

the pattern of correlations with 2-month QLS scores suggests the LP indices may improve predictive power of functional outcome assessment over conventional CVLT administration. Taken together, we find a consistent and common pattern of association for post-training performance and the categorical LP index, while regression residuals, pre-post difference scores, and gain scores correlate only with QLS.

Given the patterns of relationships observed across examination of test-retest reliability, interrelatedness across LP measures, and association with independent measures of learning (i.e., WCST) and functioning, several conclusions can be made about the strengths and weakness of different methods of computing LP. Because test-retest reliability was remarkably poor for gain scores, we are cautious in suggesting this computational approach as a preferred measure of LP. While test-retest reliability was relatively high for the simple pre-post difference score, this method is also not without limitations. Since difference scores have a negative correlation with pre-training scores, high difference scores are more apt to be obtained when pre-training performance is low (Linn & Slinde, 1977), an effect further confounded by possible regression to the mean (Klauer, 1993; Schoettke et al., 1993). There are also scaling issues, as difference scores do not necessarily have the same distributional characteristics as the raw pre- and post-scores from which they are derived (Schoettke et al., 1993). Although these issues are not completely mitigated by regression residuals or categorical LP classification, these methods are positively correlated with pre-training and have relatively stronger association with post-training, suggesting they are less dependent on initial ability and may better capture the capacity to benefit from instructions provided during LP assessment. Interestingly, the categorical LP index and regression residuals differed in their patterns of association with WCST and QLS validity measures, wherein the categorical LP classification correlated significantly with WCST and intake QLS, while regression residual correlated with intake and 2-month QLS, but not WCST. Therefore, computational differences between these approaches appear to determine the strength of relationship with other cognitive and functional outcome measures, a somewhat undesirable characteristic.

Considering together the test-retest stability, pattern of interrelationships with other LP scores, and correlations with validity measures, post-training performance appears to confer some advantages over other LP indices. Indeed, significance testing of the strength of correlation between LP indices and QLS ratings indicated that only the post-training score produced higher coefficients than would otherwise be obtained using a standard CVLT administration. However, it should be noted that post-training score did not produce the highest bivariate correlations with QLS, and that this finding may be influenced by the intercorrelation of pre-training with post-training score. Although the William's T2 statistic is recommended as a preferred approach for contrasting the strength of dependent correlations (Steiger, 1980), this method takes into account the intercorrelation between

predictors, and thus may be more sensitive for contrasts involving highly related predictors, such as pre-training and post-training score, and less sensitive when predictors share a zero-order correlation, as in the case of the regression residual. Furthermore, whether or not post-training performance is the preferred measure depends on exactly what the investigator wishes to capture – while the optimized post-training measure does provide an index of how well an individual can perform (which in turn is related to functioning), by not incorporating pre-testing performance, it fails to assess the *amount* of learning that has occurred. While strictly speaking, optimized scores do not index one's ability to learn (hence, do not index learning potential per se), they do provide a relatively stable index of potential performance, which, as suggested by some (Woonings et al., 2003), may be the best indicator of ability and the strongest predictor of functional outcomes.

ACKNOWLEDGMENTS

This work was supported in part by a grant from Veterans Administration Rehabilitation Research and Development (#D2356-R to J.F.).

REFERENCES

- American Psychiatric Association & Task Force on DSM-IV. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Bell, M., Bryson, G., Greig, T., Corcoran, C., & Wexler, B.E. (2001). Neurocognitive enhancement therapy with work therapy: Effects on neuropsychological test performance. *Archives of General Psychiatry*, *58*, 763–768.
- Budoff, M., & Friedman, M. (1964). "Learning Potential" as an assessment approach to the adolescent mentally retarded. *Journal of Consulting Psychology*, *28*, 434–439.
- Delahunty, A., Reeder, C., Wykes, T., Morice, R., & Newton, E. (2001). *Cognitive remediation therapy manual* (2nd ed.). London: King's College Institute of Psychiatry.
- Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (2000). *California Verbal Learning Test* (2nd ed.), *Adult version manual*. San Antonio, TX: The Psychological Corporation.
- First, M.B.S., Spitzer, R.L., Gibbon, M., Williams, J.B. (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition* (SCID-I/P, Version 2.0). New York: Biometrics Research Department, New York State Psychiatric Institute.
- Fiszdon, J.M., McClough, J.F., Silverstein, S.M., Bell, M.D., Jaramillo, J.R., & Smith, T.E. (2006). Learning potential as a predictor of readiness for psychosocial rehabilitation in schizophrenia. *Psychiatry Research*, *143*, 159–166.
- Green, M.F., Kern, R.S., Braff, D.L., & Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: Are we measuring the "right stuff"? *Schizophrenia Bulletin*, *26*, 119–136.
- Heaton, R. (1981). *The Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources.
- Heinrichs, D.W., Hanlon, T.E., & Carpenter, W.T., Jr. (1984). The Quality of Life Scale: An instrument for rating the schizophrenic deficit syndrome. *Schizophrenia Bulletin*, *10*, 388–398.
- Kay, S.R., Fiszbein, A., & Opler, L.A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, *13*, 261–276.
- Klauer, K.J. (1993). Learning potential testing: The effect of retesting. In J.H.M. Hamers, K. Sijtsma, & A.J.J.M. Ruijsenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues*. Lisse, Netherlands: Swets & Zeitlinger Publishers.
- Kurtz, M.M., Jeffrey, S.B., & Rose, J. (2010). Elementary neurocognitive function, learning potential and everyday life skills in schizophrenia: What is their relationship? *Schizophrenia Research*, *116*, 280–288.
- Kurtz, M.M., & Wexler, B.E. (2006). Differences in performance and learning proficiency on the Wisconsin Card Sorting Test in schizophrenia: Do they reflect distinct neurocognitive subtypes with distinct functional profiles? *Schizophrenia Research*, *81*, 167–171.
- Linn, R.L., & Slinde, J.A. (1977). The determination of the significance of change between pre- and post-testing periods *Review of Educational Research*, *47*, 121–150.
- Ohrmann, P., Kugel, H., Bauer, J., Siegmund, A., Kolkebeck, K., Suslow, T., et al. (2008). Learning potential on the WCST in schizophrenia is related to the neuronal integrity of the anterior cingulate cortex as measured by proton magnetic resonance spectroscopy. *Schizophrenia Research*, *106*, 156–163.
- Pedersen, A., Wiedl, K.H., & Ohrmann, P. (2009). Neurobiological correlates of learning potential in healthy subjects and in schizophrenic subjects. *Journal of Cognitive Education and Psychology*, *8*, 81–90.
- Rempfer, M., Hamera, E., Brown, C., & Bothwell, R.J. (2006). Learning proficiency on the Wisconsin Card Sorting Test in people with serious mental illness: What are the cognitive characteristics of good learners? *Schizophrenia Research*, *87*, 316–322.
- Schoettke, H., Bartram, M., & Wiedl, K.H. (1993). Psychometric implications of learning potential assessment: A typological approach. In J. Hamers, K. Sijtsma, & A. Ruijsenaars (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues*. Lisse, Netherlands: Swets & Zeitlinger Publishers.
- Sergi, M.J., Kern, R.S., Mintz, J., & Green, M.F. (2005). Learning potential and the prediction of work skill acquisition in schizophrenia. *Schizophrenia Bulletin*, *31*, 67–72.
- SPSS Statistics 17.0. *Release 17.0.0*. Chicago: SPSS Inc.
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.
- Tenhula, W.N., Kinnaman, J.E.S., & Bellack, A.S. (2007). Is learning potential associated with social skills in schizophrenia? *Journal of Rehabilitation Research & Development*, *44*, 827–835.
- Vaskinn, A., Sundet, K., Friis, S., Ueland, T., Simonsen, C., Birkenaes, A.B., et al. (2008). Can learning potential in schizophrenia be assessed with the standard CVLT-II? An exploratory study. *Scandinavian Journal of Psychology*, *49*, 179–186.
- Waldorf, M., Wiedl, K.H., & Schottke, H. (2009). On the concordance of three reliable change indexes: An analysis applying the dynamic Wisconsin Card Sorting Test. *Journal of Cognitive Education and Psychology*, *8*, 63–80.
- Watzke, S., Brieger, P., Kuss, O., Schoettke, H., & Wiedl, K.H. (2008). A longitudinal study of learning potential and rehabilitation outcome in schizophrenia. *Psychiatric Services*, *59*, 248–255.
- Watzke, S., Brieger, P., & Wiedl, K.H. (2009). Prediction of vocational rehabilitation outcome in schizophrenia: Incremental

- prognostic validity of learning potential beyond basic cognitive performance. *Journal of Cognitive Education and Psychology*, 8, 52–62.
- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: The Psychological Corporation.
- Weingartz, S., Wiedl, K.H., & Watzke, S. (2008). Dynamic assessment of executive functioning: (How) can we measure change? *Journal of Cognitive Education and Psychology*, 7, 368–387.
- Wiedl, K.H. (1999). Cognitive modifiability as a measure of readiness for rehabilitation. *Psychiatric Services*, 50, 1411–1413.
- Wiedl, K.H., Schottke, H.H., & Calero-Garcia, D. (2001). Dynamic assessment of cognitive rehabilitation potential in schizophrenic persons and in elderly persons with and without dementia. *European Journal of Psychological Assessment*, 17, 112–117.
- Wiedl, K.H., Schottke, H., Green, M.F., & Nuechterlein, K.H. (2004). Dynamic testing in schizophrenia: Does training change the construct validity of a test? *Schizophrenia Bulletin*, 30, 703–711.
- Wiedl, K.H., & Wienobst, J. (1999). Interindividual differences in cognitive remediation research with schizophrenic patients: Indicators of rehabilitation potential? *International Journal of Rehabilitation Research*, 22, 55–59.
- Woonings, F.M.J., Appelo, M.T., Kluiter, H., Slooff, C.J., & van den Bosch, R.J. (2003). Learning (potential) and social functioning in schizophrenia. *Schizophrenia Research*, 59, 287–296.