Short-form *Philadelphia Naming Test*: Rationale and Empirical Evaluation

Introduction

The Philadelphia Naming Test (PNT) has been used extensively in psycholinguistic, computational, and clinical research (e.g., Abel et al., 2009; Dell et al., 2007; Martin et al., 1994; Shapiro & Caramazza, 2003) on account of its favorable psychometric properties. Chief among these is its considerable length (175 items) and its detailed system for coding accuracy and errors. In many respects, the PNT also holds promise as a clinical instrument. Naming is a complex cognitive process involving the interplay of semantic, lexical, and phonological language systems; and naming impairments are commonly diagnosed, and commonly treated, in the aphasia clinic. Different from the *Boston Naming Test* (Kaplan et al., 1983), item difficulty on the PNT does not increase across the set. The lexical properties of targets (e.g., frequency and familiarity) are such that all items are likely to fall within participants' pre-morbid naming vocabulary. In a stepwise regression analysis, aphasia severity, measured by the *Western Aphasia Battery* Aphasia Quotient (WAB AQ; Kertesz, 1982), was a strong, significant predictor of PNT accuracy (n = 121; Std. beta = .88; R^2 = .77); and, importantly, age, education, and other demographic variables made no independent contribution to the model (all betas < .1).

Given its favorable target properties, and the abundance of publically available findings from participants with aphasia (see: www.mappd.org), clinical evaluations could benefit from the empirical backing of this assessment tool. However, the PNT is an impractical tool for the clinic. The large number of items confers reliability, but it also means that the test can take up to an hour or more to administer. We sought to maintain the favorable properties of the PNT while reducing the set to a more clinically manageable 30 items, the results of which could be compared directly to the large body of PNT research findings to characterize individual performance levels. In addition, we sought to create two unique sets of items, matched identically to the PNT's properties, for the purpose of measuring spontaneous or treatment-related change.

Methods

Two forms with 30 different items (PNT30-A and PNT30-B) were generated from the PNT. Each matches the PNT's distribution of target frequency, length, and semantic category exemplars. Item selection was further constrained to preserve severity-by-error type interactions from a PNT study of 94 patients (Schwartz et al., 2006). Items were excluded that elicited high rates of omissions in patients or had questionable visual clarity or name agreement according to 20 healthy participants, age-matched to the patients.

Performance on the PNT30 was evaluated in a sample of 25 individuals with chronic aphasia secondary to left hemisphere stroke. On the full PNT, administered at least 6 months earlier, accuracy scores from these individuals spanned evenly over the full range. In the short-form study, they each performed both short forms within one week, and they performed the full PNT twice, also within one week, with a month intervening between the short- and full-form administrations. Half the participants performed the

short forms before the PNT; and the order of the short forms, A and B, was also counterbalanced.

Results

To accommodate the fact that accuracy scores are dichotomous (right/wrong) and scale between 0% and 100%, the scores were transformed to the empirical logit for the purpose of calculating correlations (McCullagh & Nelder, 1989). The full PNT test-retest correlation was nearly perfect (r = .99, Figure 1a). The correlations of each short form with the first full PNT were almost as high (r = .93 and .98 for PNT30-A and -B, respectively; see Figure 1b and c). These data justify the translation of PNT30 scores to PNT equivalent scores for the derivation of percentile norms based on the abundant PNT research data. Specifically, a simple table can be derived that enables a clinician to see how an individual's PNT30 score translates to a PNT score, and what the percentile rank of that score is, relative to archived patient norms.

Further analysis revealed that each short form was highly consistent with the other (r = .93; See figure 1d). This justifies the use of PNT30-A and -B as alternative forms for measuring change in experimental and clinical settings.

An important consideration in measuring change is the inherent variability of the test. We explored the test-retest variance in the PNT to estimate an upper limit on the expected difference between the short forms. PNT test-retest difference scores had an approximately normal distribution centered near zero (Mn. = 0.01% or 2 items; Std. dev. = 0.04% or 7 items). Difference scores for PNT30-A vs. PNT30-B also approximated a zero-centered normal distribution. Not surprisingly, given the many fewer test items, the variability was higher (Mn. = -0.02% or -0.6 items; Std. dev. = 11.5% or 3.5 items). Using this error distribution to quantify the variability between the short forms, we can calculate the likelihood that an observed difference is due to random variation. By setting a threshold for an acceptable likelihood of chance findings (e.g., 1 standard deviation), we can set a target for clinically significant improvement at a given level of deficit.

Conclusions

The aim of this project is to translate research findings from the PNT into a clinical tool for diagnosis and measurement of change. We succeeded in demonstrating the reliability of the PNT for quantifying the naming impairment; and we established that the short forms produce comparable measures of performance. These findings, along with the archived research data, make possible the construction of simple look-up tables that will enable easy determination of percentile ranks for any given level of performance, as well as clinically or statistically significant indicators of change. We hope that clinicians will find the PNT30 a valuable addition to their toolbox.

References

Abel, S., Huber, W., & Dell, G. S. (2009). Connectionist diagnosis of lexical disorders in aphasia. *Aphasiology, 23*(11), 1-26.

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language, 56*(4), 490-520.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). The Boston Naming Test. Philadelphia: Lea & Febiger.

Kertesz, A. (1982). The Western Aphasia Battery. New York: Grune and Stratton.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia – Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language, 47*(4), 609-660.

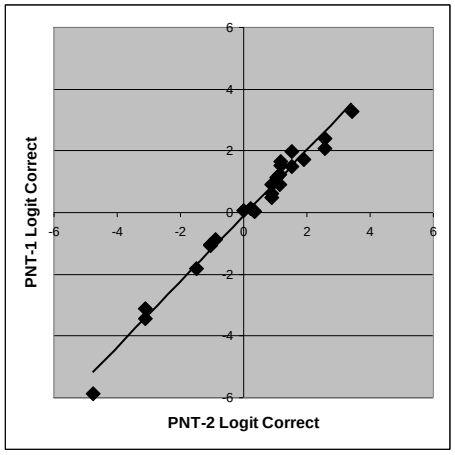
McCullagh, P., & Nelder, J. (1989). Generalized linear models. London: Chapman and Hall.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language, 54*(2), 228-264.

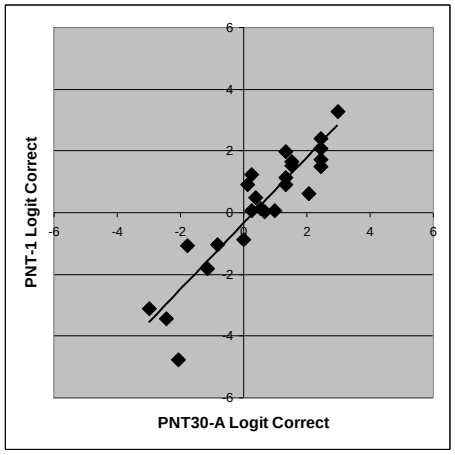
Shapiro, K., & Caramazza, A. (2003). Grammatical processing of nouns and verbs in left frontal cortex? *Neuropsychologia, 41*(9), 1189-1198.

Figure 1. Scatterplots of linear regressions comparing: (a) first vs. second full PNTs (test-retest); (b) PNT30-A with the first full PNT; (c) PNT30-B with the first full PNT; and (d) PNT30-A with PNT30-B. In all four analyses, accuracy scores were transformed to the empirical logit, and it is these transformed scores that are plotted. Model results are as follows: (a) y=1.07x-0.13; R^2=.98; (b) y=1.07x-0.37; R^2=.84; (c) y=0.89x-0.34; R^2=.86; (d) y=1.12x+0.02; R^2=.85

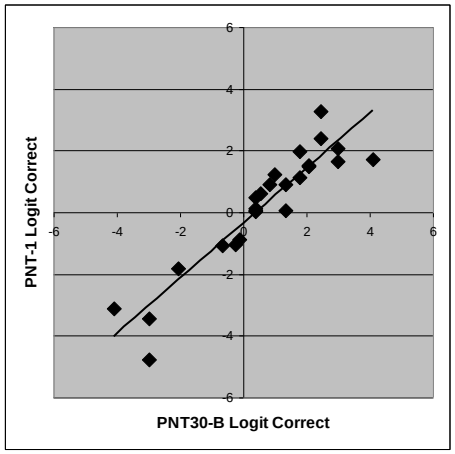
a)



b)



c)



d)