Reliability and validity of scores from the Singer-Loomis Type Deployment Inventory

Randolph C. Arnau, David H. Rosen & Bruce Thompson, College Station, Texas

Abstract: This paper focuses on the Singer-Loomis Type Deployment Inventory (SL-TDI), which is an alternative to the Myers-Briggs Type Indicator (MBTI). The SL-TDI utilizes a continuous, non-forced choice format and therefore is a more accurate representation of Jung’s personality theory of psychological types. The purpose of the study reported here is to evaluate the reliability and validity of scores from the SL-TDI. Specifically, the goals were to (a) provide estimates of the internal consistency of SL-TDI scores; (b) evaluate the divergent validity of SL-TDI scores by examination of their relationships with the scores on a social desirability responding measure; and (c) examine the test-retest stability of scores from the SL-TDI. Strong support was found for both the reliability and validity of SL-TDI scores.

Key words: Jungian personality, psychometrics, reliability, Singer-Loomis Type Deployment Inventory, validity.

Reliability and validity of scores from the Singer-Loomis Type Deployment Inventory

Jung’s (1971) theory of personality is probably the aspect of his work that has gained the most widespread acceptance in mainstream counselling and psychology. In fact, the main instrument for the measurement of Jung’s personality constructs, the Myers-Briggs Type Indicator (MBTI; Myers 1975), may well be the most popular commercially available personality instrument, with between 1.5 and 2 million persons completing it each year (Jackson, Parker & Dipboye 1996).

Although the MBTI is the most widely used instrument, there have also been serious criticisms of the instrument. For example, paired articles debating related measurement issues have appeared in the Journal of Counseling and Development (Carlson 1989; Healy 1989) and also in Measurement and Evaluation in Counseling and Development (McCaulley 1991; Merenda
1991). As noted by Arnau, Thompson and Rosen (1999), there have been two main criticisms of the MBTI. One criticism is the use of a forced-choice response format (see Cowan 1989; Garden 1991), which produces spurious negative correlations among items (Kerlinger 1986). Another criticism of the MBTI is the bipolarity assumption inherent in the scoring system which yields dichotomous types rather than continuous scores (Loomis & Singer 1980). In other words, the MBTI assumes, a priori, that an individual is one particular type or another, rather than allowing varying development on both attitudes and all functions.

Metzner and Mahlberg (1981) have suggested that ‘the rigid dichotomizing inherent within the functional typology … has served as a kind of conceptual strait-jacket, inhibiting growth and development of the model’ (p. 33). Another Jungian personality instrument, the Grey-Wheelwright Jungian Type Survey (GWJTS; Wheelwright, Wheelwright, & Buehler 1964), is also not free of criticism, as it utilizes a forced-choice format. However, it should also be noted that some Jungian measures, such as the Personal Preferences Self-Description Questionnaire (PPSDQ; Mittag 1999; Vacha-Haase & Thompson 1999) yield continuous rather than dichotomous type scores.

An examination of Jung’s (1971) theory of personality development reveals that these criticisms are definitely warranted. Although Jung did often discuss natural preferences for opposing attitudes and functions over their opposites, he did not assume that preferences are always clear or definitive. For example, Jung (1971) noted,

... there must be two fundamentally different general attitudes which would divide human beings into two groups – provided the whole of humanity consisted of highly differentiated individuals. Since this is obviously not the case, one can only say that this difference of attitude becomes plainly observable only when we are confronted with a comparatively well-differentiated personality ...

(p. 549; emphasis added)

In addition, not only did Jung not assert that preferences are always definitive, but also his theory clearly leaves room for development of functions over the lifespan. Indeed, Jung postulated the development of relatively less developed functions as an integral part of the individuation process. Therefore, theoretically speaking, a system of measurement for Jungian types should allow for varying levels of development of the functions and attitudes, rather than simply classify a person as one type or another (that is, Extravert vs. Introvert).

Although a number of empirical research studies have been conducted in an attempt to validate the bipolarity assumption, results have been somewhat mixed and the methods used have not always been optimal for the research question (for a review, see Arnau & Rosen 2000). However, one study has been conducted using a statistical procedure explicitly designed to test whether or not a construct manifests itself as a latent dimension versus a discrete class. That study provided some evidence that the Jungian attitudes and functions
are actually dimensional constructs (Arnau, Green, Gleaves, Rosen & Melancon 1998) and not categorical types, thus lending more weight to the criticism of the dichotomization procedure inherent in the MBTI scoring system.

Singer-Loomis Type Deployment Inventory

History and theoretical rationale

The Singer-Loomis Type Deployment Inventory (SL-TDI; Singer, Loomis, Kirkhart & Kirkhart 1996a, 1996b) is a recent revision of the Singer-Loomis Inventory of Personality (SLIP; Singer & Loomis 1984). The SL-TDI was developed as an alternative to the MBTI and similar measures. The SL-TDI is structured in line with the hypothesis that people are not one particular type, but rather, an individual possesses varying levels of each attitude and function. In other words, the Jungian attitudes and functions are theorized, from the standpoint of the SL-TDI, to be independent, dimensional constructs. Therefore, the response format of the SL-TDI is continuous and not forced-choice. In this manner, the SL-TDI is similar to the PPSDQ, as noted previously.

The decision to develop the SLIP came after an empirical investigation (Loomis & Singer 1980) of the bipolarity assumption. Specifically, Loomis and Singer rewrote the items on the MBTI and the GWJTS such that the instruments measured each attitude and function independently with a non-forced choice format. In other words, the formats were revised such that the test-taker could endorse varying degrees of each attitude and function, rather than being forced to choose between one and the other for each question. Loomis and Singer administered both the revised and original forms of the GWJTS to 120 participants and the revised and original forms of the MBTI to 79 participants.

To address the question of whether the bipolarity observed in the MBTI is an artefact of the forced-choice format, Loomis and Singer (1980) examined two pieces of information. First, they examined the agreement between the original and revised forms in classification of superior or dominant (most highly developed) function. Second, they examined the number of participants for whom the revised, non-forced choice form showed an inferior (least highly developed) function that was not the theoretical opposite of the measured superior function (which never happens with the original MBTI due to the forced-choice format). For the GWJTS, they found that 72% of the participants exhibited a change in measured superior function across the two test forms. In addition, when using the non-forced choice GWJTS form, 55% of the participants had an inferior function that was not the theoretically bipolar opposite of the superior function.

For the experimental non-forced choice MBTI, 46% had a superior function change and 29% had an inferior function that was not the bipolar opposite of the superior function. Based on these results, Loomis and Singer (1980) concluded that the bipolarity assumption is not valid, and that the bipolarity
observed with the conventional MBTI is largely an artefact of the forced-choice nature of the test.

Another assumption underlying the structure of the SL-TDI (unlike the structure of either the MBTI or the GWJTS) involves the two basic attitudes (that is, Extraversion-Introversion). The SL-TDI posits the two Jungian attitudes to be inseparable from the functions and assumes instead that an individual uses the functions in either an introverted or extraverted manner. Thus, for example, the MBTI, the GWJTS, and the PPSDQ yield scores on both Introversion and Thinking, but the SL-TDI yields scores such as Extraverted Thinking and Introverted Thinking.

Test format

The SL-TDI consists of 20 different hypothetical situations, each followed by a list of eight possible reactions to the situation. Each reaction corresponds to a combination of an introverted or extraverted orientation with each of the four functions (for example, Extraverted Thinking, Introverted Thinking). The respondent indicates on a five-point Likert scale how often he or she would make that response (1 = never, 5 = always).

When the SLIP was revised in 1996 (at which time the name was changed to the SL-TDI), the wording was changed for some of the hypothetical situations and possible responses. Additionally, some hypothetical situations were deleted, while some new situations were added. As Kirkhart and Kirkhart (1998) explained, the primary objective of this revision was to update the situations and possible responses such that they would better ‘reflect contemporary experience and broaden applicability to a wider spectrum of socio-economic groups’ (p. 5).

Scoring

Scores for the SL-TDI for each scale are derived by summing the responses to each item tapping into a given scale. Doing so for each of the measured scales yields eight scores (4 functions each combined with 2 attitudes). These scores can be, in turn, aggregated to obtain scores for the overall attitudes and functions. For instance, adding the scores for Extraverted Thinking and Introverted Thinking yields an aggregated score for overall Thinking. Likewise, adding scores from all 4 scores involving Introversion (that is, Introverted Thinking, Introverted Feeling, Introverted Sensation, Introverted Intuition) yields an aggregated score for overall Introversion.

In addition to summing responses to obtain raw scale scores, the manual recommends deriving percentage scores by dividing an individual’s raw scores from each scale by the total of all the individual’s responses on all scales (Singer & Loomis 1984, p. 4). This, in effect, yields scores that reflect the percentage development of the functions for a given individual. The authors of the
SL-TDI recommended this procedure to reduce the effects of response sets, which are tendencies for an individual to either respond with mostly high numbers or mostly low numbers. The disadvantage of this scoring procedure is that it leads to ipsatised scores, which means the frame of reference becomes the individual rather than some normative sample. In other words, when scores are ipsatised, the level of each score is expressed in relation to the level of the individual’s other scores (Anastasi 1982, p. 517). In addition, invoking ipsative scoring inherently yields spurious negative correlations among items (Kerlinger 1986, p. 463). Because of these properties of ipsatised scores, when the SL-TDI is used in research involving comparisons of scores across individuals, the raw scores, rather than the ipsatised, percentage scores should be used.

Very little research has been conducted with the new version of the SL-TDI. The technical manual reports internal consistency reliability of scores from the instrument ranging from .67 to .90 (Kirkhart & Kirkhart 1998, p. 7), which can be considered acceptable to excellent. Only one published study has reported on the reliability and validity of the new version of the SL-TDI. In that study, the internal consistency reliability of data from the SL-TDI ranged from .64 to .90 (Arnau et al. 1999) which are also considered acceptable to excellent. In addition, evidence for the validity of SL-TDI scores came from moderate bivariate and multivariate correlations with another alternative measure of Jungian personality, the PPSDQ (Thompson 1996; see also Kier, Melancon, & Thompson 1998; Thompson & Arnau 1998; Thompson & Melancon 1995).

Validity of SL-TDI scores has also been demonstrated through predicted correlations with similar constructs from the Five-Factor Model (FFM) of personality. The FFM conception of personality is currently one of the most well accepted models of personality constructs (Goldberg 1993; McCrae & Costa 1989). The FFM represents a culmination of extensive factor analyses of personality trait descriptors that have consistently provided evidence for five broad factors of personality (for example, Digman 1990; McCrae & Costa 1987; McCrae & John 1992). These five factors are commonly referred to as Extraversion, Neuroticism, Agreeableness, Openness to Experience, and Conscientiousness. The following predicted relationships between SL-TDI and FFM constructs were reported by Arnau et al. (1999): SL-TDI Extraversion and FFM Extraversion (r = .36), SL-TDI Introversion and FFM Neuroticism (r = .31), SL-TDI Intuition and Neuroticism (r = .31), SL-TDI Thinking and FFM Conscientiousness (r = .31), and SL-TDI Feeling and FFM Agreeableness (r = .14; all p’s < .05).

Because internal consistency of responses to an instrument is so important, one purpose of the present study was to provide further estimates of the internal consistency of data from the SL-TDI. Another important type of reliability is temporal stability, or test-retest reliability, which can be thought of as the extent to which a person’s scores on an instrument remain constant from one occasion to another (DeVellis 1991, p. 37). When an instrument is
measuring constructs that are considered to be fairly stable over time, such as personality traits, temporal stability is extremely important. To date, no published study has reported the temporal stability of scores from the SL-TDI. Therefore, another purpose of the present study was to examine test-retest reliability of SL-TDI scores.

Even if scores from an instrument are reliable, this does not guarantee that they are valid, because reliability is necessary, but not sufficient for validity (Nunnally & Bernstein 1994). In other words, reliability is necessary to show that scores are a measure of some construct, but does not necessarily mean that the constructs measured are the ones the instrument is purported to measure. As mentioned previously, validity for SL-TDI scores has been demonstrated through correlations with scores from another Jungian personality instrument as well as through correlations with scores on a measure of the FFM constructs. This type of validity, where scores are correlated with scores from other instruments measuring the same or similar constructs, is called concurrent validity. However, another important aspect of validity is that scale scores do not correlate with other constructs with which they should not be theoretically related. This type of validity is commonly called divergent validity. Because no published studies have examined divergent validity of SL-TDI scores, another purpose of the present study was to provide evidence for the divergent validity of SL-TDI scores. This was accomplished through an examination of the relationship between SL-TDI scores and scores from a measure of socially desirable responding (Reynolds 1982).

In summary, the purposes of the present study were to (a) provide additional estimates of the internal consistency of SL-TDI scores, (b) evaluate the divergent validity of SL-TDI scores by examination of their relationships with the scores on a social desirability responding measure, and (c) examine the test-retest stability of scores from the SL-TDI. Neither (b) nor (c) have been reported previously.

Method

Participants

A total of 165 college students participated in this study for partial fulfilment of course requirements for Introductory Psychology. Study participants included 57 males (34.5%) and 108 females (65.5%), and most were either in their first year (51.5%) or second year (31.5%) of college. The ethnicity of the sample was primarily Caucasian (87.9%), but also included Hispanic (7.3%), African-American (2.4%), Asian-American (1.8%) and Polynesian (0.6%) participants. The sample included participants with a wide range of major areas of study, the most prevalent being Business (33.9%), Science (14.5%), Liberal Arts (13.9%), General Studies (12.7%), and Education (7.3%).
Procedure

After signing informed consent forms, participants completed a packet of questionnaires, including a brief demographics questionnaire, the SL-TDI and a short form of the Marlowe-Crown scale of socially desirable responding. Participants returned approximately two weeks later and completed the SL-TDI again. Four participants were excluded from analyses due to missing responses at time 1, and only 150 participants completed questionnaires at time 2 with 5 participants excluded from analyses due to missing responses. Therefore, analyses for time 1 are based on 161 participants, whereas analyses which include time 2 data are based on 145 participants.

Instruments

SL-TDI

This instrument has already been described above. However, one additional comment on our use of this instrument is warranted. As mentioned previously, the SL-TDI authors recommended calculating percentage scores for interpretation, whereas the raw scores generally should be used for research, because ipsatisation does not allow for comparisons across individuals. Nevertheless, due to the potential importance of both raw and ipsatised scores from the SL-TDI, the test-retest reliability of both the ipsatised and raw scores were evaluated in the present study.

Marlowe-Crown Social Desirability Scale (Short Form)

The social desirability scale used was a 13-item short form of the Marlowe-Crown Social Desirability Scale, described by Reynolds (1982). Reynolds found scores on this short form to have acceptable internal consistency. In addition, he found that the total scores were highly correlated with the full length Marlowe-Crown instrument ($r = .93$), and moderately correlated ($r = .41$) with the Edwards Social Desirability Scale (Edwards 1957).

Results

Internal consistency

The internal consistency of scores from each scale was calculated using Chronbach’s alpha. Table 1 presents the alpha coefficients for scores on each of the SL-TDI scales. Alpha coefficients are presented separately for time 1 and time 2 administrations. The alpha coefficients ranged from .65 to .91.
Divergent validity

The divergent validity coefficients for the SL-TDI scale scores are presented in Table 2. It was predicted that the SL-TDI scale scores would be uncorrelated with the Marlowe-Crown social desirability scores. The correlations ranged from 0.0 to .17.

Temporal stability

Raw scores. The correlations across time points for raw scale scores are presented in Table 3 (lowest aggregation level) and Table 4 (middle and highest aggregation levels), with test-retest coefficients in bold-faced type. As can be seen in Tables 3 and 4, reliability coefficients for raw scale scores at the lowest level of aggregation ranged from .62 to .79, while those for the two higher level of aggregation ranged from .67 to .80. It should also be noted that there

Table 1: Cronbach’s Alpha Coefficients

<table>
<thead>
<tr>
<th>Scale</th>
<th>Time</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Lowest Aggregation Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introverted Thinking</td>
<td>.72</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Extraverted Thinking</td>
<td>.74</td>
<td>.75</td>
<td></td>
</tr>
<tr>
<td>Introverted Feeling</td>
<td>.65</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td>Extraverted Feeling</td>
<td>.69</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Introverted Sensation</td>
<td>.71</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Extraverted Sensation</td>
<td>.72</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>Introverted Intuition</td>
<td>.74</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>Extraverted Intuition</td>
<td>.73</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Middle Aggregation Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thinking</td>
<td>.85</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Feeling</td>
<td>.79</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Sensation</td>
<td>.83</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Intuition</td>
<td>.84</td>
<td>.85</td>
<td></td>
</tr>
<tr>
<td>Highest Aggregation Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introversion</td>
<td>.90</td>
<td>.91</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>.91</td>
<td>.91</td>
<td></td>
</tr>
</tbody>
</table>

Note. Due to the way the SL-TDI is scored, items are aggregated at three levels; at the highest level, there are only two scores, both of which involve more items than scores aggregated at either of the two lower levels.
are also noteworthy correlation coefficients among other scales across time points. The implications and possible reasons for this finding will be examined in the discussion.

Ipsatised scores. The correlations across time points for the ipsatised scale scores are presented in Table 5 (lowest aggregation level) and Table 6 (middle and highest aggregation levels), with the test-retest coefficients in bold-faced type.

As can be seen in the Tables 5 and 6, reliability coefficients for ipsatised scale scores at the lowest level of aggregation ranged from .32 to .52, while those for the two higher levels of aggregation ranged from .28 to .46.

Discussion

The purposes of the present study were to evaluate several psychometric properties of SL-TDI scores: internal consistency reliability, divergent validity, and test-retest reliability. The internal consistency of data from the SL-TDI in
Table 3: Raw Scale Correlation Coefficients Across Times
For Lowest Aggregation Level

<table>
<thead>
<tr>
<th>Scales</th>
<th>IS2</th>
<th>IN2</th>
<th>IT2</th>
<th>IF2</th>
<th>ES2</th>
<th>EN2</th>
<th>ET2</th>
<th>EF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1</td>
<td>.71</td>
<td>.65</td>
<td>.63</td>
<td>.61</td>
<td>.54</td>
<td>.60</td>
<td>.55</td>
<td>.43</td>
</tr>
<tr>
<td>IN1</td>
<td>.60</td>
<td>.74</td>
<td>.50</td>
<td>.53</td>
<td>.43</td>
<td>.53</td>
<td>.38</td>
<td>.31</td>
</tr>
<tr>
<td>IT1</td>
<td>.58</td>
<td>.54</td>
<td>.79</td>
<td>.58</td>
<td>.55</td>
<td>.60</td>
<td>.67</td>
<td>.47</td>
</tr>
<tr>
<td>IF1</td>
<td>.48</td>
<td>.54</td>
<td>.48</td>
<td>.62</td>
<td>.41</td>
<td>.53</td>
<td>.41</td>
<td>.38</td>
</tr>
<tr>
<td>ES1</td>
<td>.57</td>
<td>.49</td>
<td>.61</td>
<td>.48</td>
<td>.74</td>
<td>.56</td>
<td>.58</td>
<td>.50</td>
</tr>
<tr>
<td>EN1</td>
<td>.51</td>
<td>.59</td>
<td>.60</td>
<td>.52</td>
<td>.48</td>
<td>.77</td>
<td>.55</td>
<td>.50</td>
</tr>
<tr>
<td>ET1</td>
<td>.54</td>
<td>.45</td>
<td>.71</td>
<td>.54</td>
<td>.59</td>
<td>.58</td>
<td>.70</td>
<td>.50</td>
</tr>
<tr>
<td>EF1</td>
<td>.36</td>
<td>.43</td>
<td>.46</td>
<td>.43</td>
<td>.45</td>
<td>.55</td>
<td>.45</td>
<td>.69</td>
</tr>
</tbody>
</table>

Note. N = 145. Labels ending in “1” indicate time 1 scores and labels ending in “2” indicate time 2 scores. ES = Extraverted Sensation, EN = Extraverted Intuition, ET = Extraverted Thinking, EF = Extraverted Feeling, IS = Introverted Sensation, IN = Introverted Intuition, IT = Introverted Thinking. All correlations are statistically significant at the p < .01 level. Test-retest reliability coefficients are in bold.

Table 4: Raw Scale Correlation Coefficients Across Times
For Middle and Highest Aggregation Levels

<table>
<thead>
<tr>
<th>Scales</th>
<th>Extraversion</th>
<th>Introversion</th>
<th>Sensation</th>
<th>Intuition</th>
<th>Thinking</th>
<th>Feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Extraversion 1</td>
<td>.77</td>
<td>.69</td>
<td>.66</td>
<td>.69</td>
<td>.71</td>
<td>.68</td>
</tr>
<tr>
<td>Introversion 1</td>
<td>.65</td>
<td>.79</td>
<td>.66</td>
<td>.73</td>
<td>.67</td>
<td>.63</td>
</tr>
<tr>
<td>Sensation 1</td>
<td>.71</td>
<td>.75</td>
<td>.75</td>
<td>.67</td>
<td>.68</td>
<td>.62</td>
</tr>
<tr>
<td>Intuition 1</td>
<td>.63</td>
<td>.72</td>
<td>.59</td>
<td>.77</td>
<td>.59</td>
<td>.57</td>
</tr>
<tr>
<td>Thinking 1</td>
<td>.72</td>
<td>.72</td>
<td>.64</td>
<td>.62</td>
<td>.80</td>
<td>.62</td>
</tr>
<tr>
<td>Feeling 1</td>
<td>.63</td>
<td>.60</td>
<td>.50</td>
<td>.61</td>
<td>.52</td>
<td>.67</td>
</tr>
</tbody>
</table>

Note. N = 145. Labels ending in “1” indicate time 1 scores and labels ending in “2” indicate time 2 scores. All correlations are statistically significant at the p < .01 level. Test-retest reliability coefficients are in bold.
### Table 5: Ipsatised Scale Correlation Coefficients Across Times For Lowest Aggregation Level

<table>
<thead>
<tr>
<th>Scales</th>
<th>IS(_1)</th>
<th>IN(_1)</th>
<th>IT(_1)</th>
<th>IF(_1)</th>
<th>ES(_1)</th>
<th>EN(_1)</th>
<th>ET(_1)</th>
<th>EF(_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS(_2)</td>
<td>.40</td>
<td>.22</td>
<td>.03</td>
<td>.19</td>
<td>.03</td>
<td>-.03</td>
<td>.01</td>
<td>-.12</td>
</tr>
<tr>
<td>IN(_2)</td>
<td>.06</td>
<td>.41</td>
<td>-.21</td>
<td>-.12</td>
<td>-.27</td>
<td>-.19</td>
<td>-.36</td>
<td>-.38</td>
</tr>
<tr>
<td>IT(_2)</td>
<td>.03</td>
<td>-.07</td>
<td>.49</td>
<td>.06</td>
<td>.02</td>
<td>-.01</td>
<td>.28</td>
<td>-.06</td>
</tr>
<tr>
<td>IF(_2)</td>
<td>.00</td>
<td>.04</td>
<td>-.22</td>
<td>.32</td>
<td>-.10</td>
<td>-.04</td>
<td>-.17</td>
<td>.02</td>
</tr>
<tr>
<td>ES(_2)</td>
<td>.09</td>
<td>-.15</td>
<td>-.01</td>
<td>-.10</td>
<td>.50</td>
<td>-.08</td>
<td>.09</td>
<td>.07</td>
</tr>
<tr>
<td>EN(_2)</td>
<td>-.17</td>
<td>.08</td>
<td>-.03</td>
<td>.17</td>
<td>-.21</td>
<td>.43</td>
<td>-.04</td>
<td>-.04</td>
</tr>
<tr>
<td>ET(_2)</td>
<td>-.09</td>
<td>-.32</td>
<td>.25</td>
<td>-.10</td>
<td>.06</td>
<td>-.13</td>
<td>.33</td>
<td>-.08</td>
</tr>
<tr>
<td>EF(_2)</td>
<td>-.26</td>
<td>-.19</td>
<td>-.21</td>
<td>-.03</td>
<td>-.04</td>
<td>.04</td>
<td>-.09</td>
<td>.52</td>
</tr>
</tbody>
</table>

*Note. N = 145*. Labels ending in “1” are time 1 scores and labels ending in “2” are time 2 scores. ES = Extraverted Sensation, EN = Extraverted Intuition, ET = Extraverted Thinking, EF = Extraverted Feeling, IS = Introverted Sensation, IN = Introverted Intuition, IT = Introverted Thinking, IF = Introverted Feeling. Correlations with an absolute value greater than .16 are statistically significant at the p < .05 level. Test-retest reliability coefficients are in **bold**.

### Table 6: Ipsatised Scale Correlation Coefficients Across Times For Middle and Highest Aggregation Levels

<table>
<thead>
<tr>
<th>Scales</th>
<th>Extraversion</th>
<th>Introversion</th>
<th>Sensation</th>
<th>Intuition</th>
<th>Thinking</th>
<th>Feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.28</td>
<td>-.30</td>
<td>-.04</td>
<td>-.13</td>
<td>.08</td>
<td>.07</td>
</tr>
<tr>
<td>Introversion</td>
<td>-.28</td>
<td>.30</td>
<td>.04</td>
<td>.13</td>
<td>-.08</td>
<td>-.07</td>
</tr>
<tr>
<td>Sensation</td>
<td>.13</td>
<td>.15</td>
<td>.43</td>
<td>-.03</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>Intuition</td>
<td>-.28</td>
<td>-.02</td>
<td>-.24</td>
<td>.31</td>
<td>-.26</td>
<td>-.31</td>
</tr>
<tr>
<td>Thinking</td>
<td>.09</td>
<td>.04</td>
<td>.01</td>
<td>-.20</td>
<td>.46</td>
<td>-.07</td>
</tr>
<tr>
<td>Feeling</td>
<td>.06</td>
<td>-.14</td>
<td>-.15</td>
<td>-.07</td>
<td>-.25</td>
<td>.35</td>
</tr>
</tbody>
</table>

*Note. N = 145*. Labels ending in “1” indicate time 1 scores and labels ending in “2” indicate time 2 scores. Correlations with an absolute value greater than .16 are statistically significant at the p < .05 level. Test-retest reliability coefficients are in **bold**.
this study ranged from .65 to .91, which are considered acceptable to excellent (DeVellis 1991; Nunnally & Bernstein 1994). The internal consistencies of the shorter scales (lowest level of aggregation) were somewhat smaller than those of the longer scales. However, this is to be expected, because coefficient alpha is partially dependent on item variances, which can be affected by the number of items. Therefore, one might expect the reliability of scores from these shorter scales to be somewhat lower. Nevertheless, despite the relatively lower reliability of scores from these shorter scales, the internal consistencies are in the acceptable range.

Another noteworthy finding in the present study was the stability of the internal consistency estimates. As seen in Table 1, the reliabilities of scale scores at time 2 were highly consistent with the reliabilities at time 1. In addition, these internal consistencies are nearly identical to the alpha’s ranging from .64 to .90, reported by Arnau et al. (1999).

The second objective of the present study, to provide evidence for the divergent validity of scores from the SL-TDI, was accomplished through an examination of the relationship between SL-TDI scores and socially desirable responding. As socially desirable responding is theoretically unrelated to Jungian personality constructs, the SL-TDI scores were hypothesized to have little or no correlation with scores from the measure of socially desirable responding. As seen in Table 2, this was confirmed. Most correlations were below .10, and none were statistically significant. Even the highest correlation (between social desirability and Extraverted Feeling) of .17 involved a negligible effect, with social desirability explaining only 3% of the variance in Extraverted Feeling ($r^2 = .17^2 = 3.0\%$). These findings lend strong support to the divergent validity of scores from the SL-TDI.

The third objective was to examine the temporal stability (test-retest reliability) of SL-TDI scores. Because the SL-TDI manual recommends calculating relative percentage scores, the temporal stability of both the raw scores and the ipsatised percentage scores were examined. As seen in Tables 3 and 4, the raw scale scores demonstrated acceptable temporal stability across the two-week period. Conversely, the ipsatised scale scores were not very stable across the two-week period, as evidence by the low test-retest correlations.

An interesting finding was the noteworthy correlations between scales other than like scales. These correlations consist of those found in the off-diagonal elements of Tables 3 and 4. Because such high correlations involving non-identical scales is often not expected across time points, a further discussion of this finding is warranted. One important point to note about these correlations is that they are all smaller than the test-retest reliability coefficients. In other words, any given scale at time 1 correlated more highly with itself at time 2 than it did with any other scale at time 2. The only exception to this was the reliability of the Extraverted Thinking scale ($r = .70$), which was slightly lower than the correlation between Extraverted Thinking at time 1 with Introverted Thinking at time 2 ($r = .71$).
Another important point to note about these correlations is that they are consistent with the fact that these scales are, indeed, related. In other words, these inter-scale correlations across time points is consistent with the fact that these scales are also correlated within a given time point. Such correlations between scales might be expected, given that many of the SL-TDI scales are tapping into similar constructs. Although it is true that each separate scale is posited to measure unique constructs, they are also theoretically similar in some regards, due to the combination of the functions with the attitudes. For example, all of the introversion scales (IS, IN, IT, IF) are posited to be measuring different functions, but different functions all used in an introverted manner. Likewise, all of the extraverted scales are posited to measure different functions used in an extraverted manner.

Similarly, looking at these results from a different point of view, there are also two different scales for each function, an extraverted scale and an introverted scale. Therefore, scales tapping into a given function would be thought to be somewhat related constructs, even though associated with different attitudes. For example, Extraverted Sensation might be expected to be related, but not identical to Introverted Sensation. In sum, there seems to be some expected overlap between SL-TDI scales, but, importantly, these correlations across time points tend to be substantially smaller than each scale’s respective test-retest reliability coefficients.

It should be noted here that there are also some correlations on the off-diagonal that are actually higher than the test-retest correlations (on the diagonal). However, this situation is not problematic because, as can be seen in Table 3 and Table 4, the test-retest correlation of any given variable is consistently higher than the correlations of that variable with any other given variable. For example, the test-retest correlation of IS, found on the diagonal at the upper left-hand corner of Table 3, is .71. If one scans down the first column as well as across the first row of the table, it can be seen that none of the correlations of IS with any of the other scales at time 2 (across the row) or time 1 (down the column) are greater than the test-retest correlation of .71.

The analysis of the ipsatised percentage scores yielded a different picture than that of the raw scores. As can been seen in Tables 5 and 6, the off-diagonal correlations in this case were much smaller than those found with the raw score analyses, and less than half were statistically significant. Although this would seem to be a highly desirable characteristic of the ipsatised scores, this finding should be interpreted with extreme caution, because by the very nature of the ipsatisation procedure, spurious negative correlations tend to be induced (Kerlinger 1986). Another potential problem with the ipsatised scores is the relatively low temporal stability of these scores, as evidence by the test-retest reliability coefficients presented in Tables 5 and 6. Even the reliability of the longer ipsatised scales (see Table 6) is low. The conclusion from this analysis is that whereas the raw scale scores are fairly stable (at least over a two-week period), the ipsatised percentage scores are less stable. To reiterate, the practical
implications of this finding is that in any type of research or clinical setting where one is interested in development over time, the ipsatised percentage scores should not be used. Rather, the raw scale scores should be utilized.

Speculation regarding why the ipsatised scores are relatively less stable over time appears warranted. One reason could be the purported fact that the SL-TDI measures both state and trait influences on personality (Singer & Loomis 1984). If this is true, then it may be the case that the relative percentage activation of the functions (percentage scores) is more reflective of the fluid dynamics of state influences, while the raw scores are more reflective of overall development. If such were the case, the raw scores would be expected to be more stable over time whereas the percentage scores would be less stable. Therefore, the percentage scores may be an indicator of the relative activation of functions within a given individual, at a given time, with certain state influences in effect at that given time. On the other hand, the raw scores may be a better indicator of an overall development of the functions, which may be more stable over time.

Therefore, it may be the case that both raw scores and percentage scores provide meaningful information. Empirical investigation of these speculations would be a fruitful area for future research. One possible avenue for research could be to study a group of individuals for whom certain environmental influences are known, by measuring the raw and percentage SL-TDI scores both before and after the known environmental event or influence. For example, a group of college students could be tested before the start of a new semester and then again right before final exams. One might expect stability in the raw scores, but a change in the percentage scores, reflecting a change in relative activation of certain functions due to the situational demands (that is, studying for final exams).

Another possible avenue for future research would be to study both state and trait changes in response to psychotherapy. It may be the case that certain functions become relatively more activated depending on certain issues or tasks that are undertaken in therapy. For example, it may be the case that psychotherapy clients engaged in extensive dreamwork or active imagination would tend to show relatively more activation of Introverted Intuition, as opposed to clients engaging in tasks that demand more logical, rational processing. If the percentage scores are actually picking up on relative activation, independent of the level of overall development of the trait, then one would expect to see differences in percentage scores for clients engaged in different types of psychotherapeutic tasks or different modalities of psychotherapy. For example, one could compare clients in Jungian analysis with clients undergoing cognitive-behavioural therapy. It would also be interesting to examine shifts in percentage scores across time within individuals. Clients in psychotherapy could be measured before treatment, and then at various times during treatment and after treatment. Besides revealing shifts in relative function or attitude activation due to the tasks of therapy, such shifts could also convey
changing circumstances in a client’s life. These kinds of shifts may be especially
important to a clinician, given that the client’s relative personality activation
at any given point in time has been shown to influence both verbal and non-
verbal communication in psychoanalysis (Schwaber 1998).

The Jungian functions as traits (raw scores), rather than states (percentage
scores), could also be studied in the context of psychotherapy. As development
and integration of the non-dominant functions is considered a part of indi-
viduation, it would be interesting to study changes in the raw scores on the
SL-TDI from the beginning to the end of psychotherapy. Utilizing Jung’s con-
cept of individuation, one would expect successful psychotherapy to lead to
greater development of the non-dominant functions, which should be reflected
in higher raw scores on the SL-TDI.

In conclusion, the present study provides additional evidence for the reliability
and validity of SL-TDI scores. It corroborates and extends the literature on this
instrument in that results from past studies of the internal consistency of SL-
TDI scores were replicated. In addition, the present study provides evidence
for the test-retest stability and divergent validity of SL-TDI scores, which has
not been previously published. These results are even more positive in light of
the fact that the SL-TDI is free from the previously mentioned criticisms of the
MBTI. It would be useful for future studies to provide more evidence for the
convergent and divergent validity using the multitrait-multimethod approach
(Campbell & Fiske 1959). For example, an other-rated SL-TDI form could be
constructed, and correlations across the self- and other-rated forms could be
examined. Although more validity studies should be conducted, it appears that
the SL-TDI holds promise as a measure of Jungian personality constructs that
is more closely rooted in Jung’s theory of personality.

Translations of abstract

This article se concentre sur le test «Singer-Loomis Type Deployment Inventory»
(SL-TDI), qui constitue une alternative au test «Myers-Briggs Type Indicator» (MBTI).
Le SL-TDI utilise un format dans lequel les choix ne sont pas forcés et sont continus;
il donne une représentation plus juste de la théorie de Jung des types psychologiques.
Le but de l’étude dont il est rendu compte ici, est d’évaluer la fiabilité et la validité
des résultats du test SL-TDI; en particulier en ce qui concerne (a) l’estimation de la
cohérence interne des résultats du SL-TDI, (b) l’évaluation des points de divergence
apparaissant quant à la validité des résultats du SL-TDI en examinant leurs corrélations
avec les résultats d’une mesure des réactions relatives à l’attrait social, et (c) l’étude de
la stabilité des résultats du SL-TDI dans le test et re-test. Des éléments solides montrant
la fiabilité et la validité des résultats du SL-TDI ont été établis.

Diese Arbeit widmet sich dem Singer-Loomis Type Deployment Inventory (SL-TDI),
der eine Alternative zum Myers-Briggs Type Indicator (MBTI) darstellt. Der SL-TDI

Questo lavoro pone la sua attenzione sul Singer-Loomis Type Deployment Inventory (SL-TDI), che è alternativo al Myers-Briggs Type Indicator (MBTI). Il SL-TDI utilizza un formato continuo, con scelta non forzata, e rappresenta quindi una versione più accurata della teoria junghiana della personalità basata sui tipi psicologici. L’obiettivo della ricerca qui riportata è quello di valutare l’affidabilità e la validità dei punteggi ottenuti dal SL-TDI. In particolare gli obiettivi sono: (a) fornire stime della coerenza interna dei punteggi del SL-TDI, (b) valutare la validità della divergenza dei punteggi del SL-TDI esaminando i loro rapporti con dei punteggi basati su una misura rispondente a una desiderabilità sociale, e (c) esaminare la stabilità dei punteggi del test-retest presi dal SL-TDI. Si è trovata una decisa conferma sia per l’affidabilità che per la validità dei punteggi.

Este trabajo se focaliza en el Inventario de Desplegamiento Tipológico de Singer-Loomis (SL-TDI), el cual es una alternativa para el Indicador de Myers-Briggs (MBTI). El SL-TDI utiliza un formato continuo y no forzado de escogencia y es por lo tanto una representación más adecuada de la teoría de los tipos psicológicos de Jung. El propósito del estudio que se reporta aquí es para evaluarla confiabilidad y validación de los puntajes del SL-TDI. Específicamente, las metas son (a) probar las estimaciones de la consistencia interna de los puntajes del SL-TDI, (b) evaluar la divergente validación de los puntajes del SL-TDI, por medio del estudio comparativo de sus relaciones en una medida deseable de respuesta social, y (c) examinar la estabilidad de los puntajes en el test-retest del SL-TDI. Se demuestra una fuerte base para la validación y confiabilidad de los puntajes del SL-TDI.

References


Acknowledgement

The authors would like to express their appreciation to Elizabeth Kirkhart and Larry Kirkhart for their valuable support and encouragement.

[MS first received May 1999, final version November 1999]