

The Leadership Development Framework and Profile: A summary of issues of validity and reliability

Herdman-Barker, E., Rooke, D., & Torbert W. 2009.

1. Origins of Harthill's Leadership Development Profile – the Washington University Sentence Completion Test (Loevinger 1970)

Harthill's Leadership Development Profile emerges from the studies and research surrounding the Loevinger Washington University Sentence Completion Test (WUSCT). The WUSCT is itself one of the most widely used and thoroughly validity-tested instruments in developmental psychometrics (Loevinger and Wessler 1970), and it is from this base, through the ongoing research and development within Harthill that the LDP has evolved. Any consideration of reliability and validity of Harthill's LDP must necessarily draw on the extensive research into Loevinger's original work. In 1978 when Torbert began his field studies and laboratory and field experiments to explore whether differences in the developmental action-logics predicted differences in managerial style and capability, he was working with the WUSCT measure itself.

Loevinger herself was adding to the work of others and then inter-acting with contemporaries to help shape the field of knowledge which is described *constructivist developmental psychology*. Perhaps most notable in this lineage is Piaget (1952) and his seminal works on the development of children, Kohlberg (1984) and his work on the psychology of moral development, Kegan (1982,1994) and his exploration of meaning making in relationship to the changing challenges of life. This list is far from exhaustive.

Working together for many years, Torbert (1987, 1991) and Cook-Greuter (1999) set out to modify the WUSCT in several ways. First, they sought to increase its face validity and usefulness as a work-related, leadership development instrument by creating and validating work-related stems, starting from Molloy (1978) "A good boss..." Harthill, with Torbert, has continued and extended this work, so that the Harthill LDP now has nine work-related stems, including stems about power, time management, and teams that replace prior stems emphasizing gender or that had the lowest correlations to overall scf ratings. This is a meaningful and unique contribution because it expands the scope of the Harthill LDP to better map management oriented topics that are at the heart of developmental action logics in organisational contexts.

In developing WUSCT (Loevinger & Wessler, 1970; Hy & Loevinger 1996) individuals engaged in a projective completion of 36 sentence stems that cover many different aspects of human concern. Thousands of completions of the WUSCT were used to create the final instrument, drawn widely from the social spectrum, even though no single sample was very large or representative of the population as a whole (Loevinger 1998). Its development has proceeded apace since with the publication of the second edition (Hy & Loevinger, 1996).

2. Rating methodology - using the sentence completion form to create a Harthill LDP

Both the WUSCT and the LDP contain 36 sentence initial stems (e.g., 'I just can't stand people who . . .'). When completed, these reflect a subject's reasoning and thinking processes and ways of relating to others. Embedded in the content and form of the sentence completions is the subject's overall way of 'mapping their world' as conceptualized by his or her developmental action-logic.

Raters match each of the 36 responses with one of the listed category titles in the scoring manual. A rating manual is used as the basis of rating each response to a sentence stem. The rater seeks an exact match or very close match. Where a match does not exist the rater uses a rule set and their judgement to rate the action logic of the each individual sentence completion.

An unusual virtue of the scoring procedure is the rating of individual sentences across forms, rather than the 36 sentences in one form in sequence. First, a single sentence stem on many measures (typically 6 to 20) is scored separately, then the second item is rated across all forms, and so on (to avoid a 'halo' effect of early item ratings on a given measure predisposing a rater to more of the same). These scores will eventually be added together to create an overall statistical score for the individual scf.

The scores within certain ranges are categorised as indicating the dominance of a particular action-logic (a profile characterised as *Expert* may, for example, include the following distribution of responses: 2 *Opportunist*, 8 *Diplomat*, 20 *Expert*, 5 *Achiever*, and 1 *Individualist*). Before looking at the aggregate scores or computing the foregoing result, however, the rater can review each entire profile, with special attention to outlying scores, and assign a rating based on the whole gestalt. Any difference between the statistical rating and the intuitive rating leads to still further discernment, or an invitation to another trained rater to rate that profile, as a further action aimed at increasing reliability. For advanced raters, individual sentence scores are identical well over .8 of the time. Most importantly they are within a single action-logic of one another over .906 of the time. Inter-rater reliability of above .8 is considered adequate, over .9 strong.

Research demonstrates that is virtually impossible to contrive a result on the WUSCT. Experiments have shown that persons almost never succeed in producing a profile at a later action-logic than their own, even after the theory underlying the scoring procedure has been explained to them (Redmore 1976).

3. Reliability and validity of the WUSCT

Since its conception in 1970, many studies have supported the validity of the WUSCT measure (Hauser 1976, Westenberg, Blasi and Cohn 1998, Cox 1974, Redmore and Waldman 1975). Theorists and studies have extensively reviewed its validity and reliability; Loevinger and Wessler (1970) report high inter-rater reliability and internal consistency on the WUSCT, and Loevinger (1976, 1998). Other studies generally report high levels of inter-rater reliability (e.g. Cox, 1974, Hoppe 1972), construct validity (Hauser 1976, Westenberg Vaillant and McCullough 1987, Hy and Bobbitt 1991) and split-half reliability (Novy & Francis, 1972; Redmore & Waldman, 1975). Westenberg et al (1998) in a study of children and adolescents support the cross-cultural and cross-age validity of the ego maturity model. External validity was researched by Rock (1975), clinical validation by Hauser (1976), Powers and Noam (1991) co-validated the WUSCT with moral reasoning.

4. Recent developments in improving the validity and reliability of the LDP

The Harthill's Leadership Development Profile has improved the WUSCT in a number of ways:

- Harthill's Leadership Development Profile adds independently validated work-related sentence stems (e.g. 'A good boss...', Molloy, 1978) to increase the measure's face validity for people in work settings. Stems about teams, time and power have been

added to replace older gendered stems (e.g. 'Men are lucky...' 'Women are lucky...') that generate responses with the lowest correlations to overall profiles.

- The most recently added six stems have proven to correlate with overall profile ratings better than the replaced stems, and also better than the average of the remaining original stems. In other words, the over-all reliability of the LDP is increasing. (As of 2007, the correlations between individual sentence stems and the whole profile ranged from .47 to .63, with an average of .55.)
- In a 2007 construct validity study of the LDP, yet to be published in detail, Reut Livne-Tarandach performed two sets of factor analyses of the 36 items in the measure – one of the 'Conventional' profiles (Achiever [Stage 6] and earlier) (n=830); and one of all 'Post-Conventional' profiles (Individualist [Stage 7] and later) (n=61). This study permits us to see whether there are any differences in the factors produced by Conventional and Post-Conventional profiles and whether such differences support the construct validity of the developmental distinctions. The results show, not only significantly different factors in the two sets of profiles, but also a significant difference in the structure of those factors. These results are consistent with theoretical predictions: the Conventional profiles generate a relatively simple mental map with eight distinct, independent clusters; in contrast, the Post-Conventional profiles generate a more complex mental map with eleven systems-oriented, interdependent clusters (more than half of the stems load on two or more factors (9 on 2, 7 on 3, and 3 on 4).
- The quality of the scoring manuals is being consistently up-graded as the data-base of respondents grows. This is particularly the case for the later action-logics, re-aligning them more closely with the Alchemist (Stage 9) and Ironist (Stage 10) constructs of Cook Greuter (1999) and Torbert's (1987 2004) work. Sheer cognitive complexity is not synonymous with the ongoing experiential research that an Alchemist exercises in his or her work, family, and leisure. This theoretical and experiential sense of the Alchemist (Stage 9) action-logic has led to a further revision of the Harthill LDP Alchemist scoring manual, including additional new criteria, and we have raised the number of Alchemist stems and the number of categories those stems must fall into for a total profile to be scored early Alchemist or full Alchemist. All this, we believe, increases the construct validity of the Alchemist designation when it is used to summarise the centre of gravity action-logic represented by a person's LDP.
- Training and development of raters is ongoing and reliability scores continue to rise. Harthill has carefully trained a number of raters over the last decade. The minimum training period is two to three years.
- In terms of the potential usefulness to executive coaches and consultants of the profile as a feedback instrument for leaders, the last decade has seen the introduction and refinement of an individual commentary of between 200 to 300 words for each profile, written by one of only three specifically trained commentary writers. This is aimed at alerting the executives and any consultant / coach working with them to themes that emerge in the structuring and content of the respondent's completions. This commentary does not impact the rating of action logic, but extends what respondents may gain developmentally from reflecting on their current action-logic rating. (For economy reasons some consultants and individuals prefer to receive profiles without commentary.)

5. External and Catalytic Validity

Whilst there was an extensive body of internal validity research about the WUSCT, there was very little external validity data available when Torbert began adapting Loevinger's work to the organisational environment and it was pragmatic, real-world validity with which he was primarily concerned.

Early on, Hauser (1976) showed statistically significant relationships between the WUSCT and interpersonal behaviour. Since then, Harthill's laboratory and field action inquiry studies have augmented the validity-testing of this approach to leadership by concentrating on researching differences in the real-world performance of people. We have found:

1. Statistically significant differences in managerial performance between Conventional and Post-Conventional action-logics on in-basket tests (Merron, Fisher & Torbert, 1987) and in an interview study (Fisher & Torbert, 1991). As theoretically predicted, Post-Conventional leaders were more likely to engage action dilemmas in inquiring, collaborative ways.
2. Another study (Torbert & Fisher, 1992) shows that voluntary participation in groups that encourage first and second-person action inquiry aimed at creating shifts in meaning making, over a two to four year period (formed and guided by a Post-Conventional action-logic practitioner) generated transformation to post-conventional action-logics among 22 of the 24 participants. By contrast, only 3 of 165 persons in a control group (who started the same MBA program at the same time as the participants) showed positive developmental transformation over the same period when retested. This finding accounts for an unusually high percentage of the variance – 91%.
3. A field study (n=281) (Torbert, 1994), during which all participants were invited to request feedback on the LDP measure's findings about them, if they so wished, found that a larger proportion at each later developmental action-logic asked for feedback, a perfect 1.0 correlation on a Spearman Rank Order test, confirming the theoretical prediction that Post-Conventional action-logics are increasingly open to (and increasingly seek out) single- and double-loop feedback and learning. Understood this way, the finding supports both construct & external validity.
4. In a further field study a complex task set for 16 teams (Torbert, 1987) showed that those teams (5 of 16) with one or more members measured at Strategist (stage 8) outperformed teams with no one measured Strategist in three ways: in terms of bottom-line effectiveness, and in terms of members' perceptions of efficient time-use and greater within-group support for own learning.
5. In an LDP study of 10 CEOs and their Lead Consultants in 10 organizations attempting to transform (Rooke & Torbert, 1998), only the five CEOs measured at the Strategist (8) action-logic reliably succeeded in generating organizational transformation (accounting for .42 of the variance at a .05 level, Spearman Rank Order test). In this case, of the four Lead Consultants, three are measured at the Strategist action-logic, one at Alchemist. As theoretically-predicted, in the five cases when the CEO is measured as at an earlier-than-Strategist action-logic, only the consultant measured at the Alchemist action-logic succeeds in supporting the two organizations that nevertheless succeeded in transforming. The 2 non-Strategist CEOs matched with Strategist Lead Consultants made some money but no

sustained transformational change in the two organizations they were involved with. And, in the case of the Diplomat CEO [the earliest-recorded action-logic among these ten CEOs], who was paired with the Alchemist action-logic Lead Consultant, the consultant resigned after the CEO and Board Chair chose not to act on the consultant's recommendation that the CEO resign or be fired. Thus, if one adds the LDP scores of each CEO/Lead Consultant duo, then the combined influence of their action-logics accounts for 59% of the variance (at a .01 level, on that same Spearman Rank Order test) in whether the organization succeeds in transforming. The small sample size and the non-parametric statistical test highlight the construct and external validity of the measure as a metric that properly distinguishes among individual cases. We know of no evidence that any other developmental measure has tested or attained this kind of field/catalytic/external validity (see appendix of Torbert & Associates, 2004).

6. Critical perspectives.

The validity of the sentence completion methodology has been criticised on two counts: (1) it is simply a matter of verbal fluency, since later action-logic respondents write significantly longer scfs: and (2) ego development's status as a 'master trait,' as proposed by Loevinger, has not been proven psychometrically. On the issue of verbal fluency, it may be countered that the correlation is inevitable, since the expression of greater conceptual complexity and creativity characteristic of the later stages requires longer responses (Loevinger & Wessler, 1970; Vaillant & McCullough, 1987). At the same time, it is clear from numerous studies that ego development, as measured by the WUSCT, entails many more facets than verbal fluency. Personality characteristics found appropriately associated with predicted developmental action-logics include rule-boundedness (in this case, a curvilinear relationship, lower in pre and post-conventional action-logics, higher at conventional action-logics), nurturance, conscientiousness, trust, tolerance, interpersonal sensitivity, psychological mindedness, creativity, moral development and a variety of measures related to mental health (Kohlberg, 1963,1964; Lorr & Manning, 1978; Vaillant, 1977; Vaillant & McCullough, 1987).

A study by Novy et. al. (1994) tested the structural validity of Loevinger's model, using structural equation modelling to examine whether ego development serves hierarchically as the source of co-variation for specific personality constructs. While not proven to be a 'master trait' by this analytic model, ego development was positively and significantly related to each of the four constructs posited by Loevinger: impulse control, interpersonal style, conscious preoccupations and cognitive style. Those wishing to review the scholarly literatures on the WUSCT can also refer to the 1993 special issue of *Psychological Inquiry* devoted to Loevinger's work, as well as the Westenberg et al.1998 book in honour of her 80th birthday, *Personality Development: Theoretical, Empirical, and Clinical Investigations of Loevinger's Conception of Ego Development*.

References

- Carlson, V. & Westenberg, P. (1998). "Cross cultural application of the WUSCT," in J. Loevinger, ed., *Technical foundations for measuring ego development*. 57-75.
- Chandler, D. & Torbert, W. (2003). Transforming inquiry and action: Interweaving 27 flavors of action research. *Journal of Action Research* 1, 2, 133-152.
- Cook-Greuter, S. (1999). "Postautonomous ego development: A study of its nature and measurement." Cambridge, MA, Harvard University Graduate School of Education, doctoral dissertation.
- Cox, L. (1974). "Prior help, ego development and helping behavior" *Child Development*. 45, 594.
- Fisher, D., and Torbert, W. (1991). Transforming Managerial Practice: Beyond the Achiever Stage, (143-173). In R.W. Woodman & W. A. Pasmore (Eds.). *Research in organization change and development*. Vol. 5. Greenwich, CT: JAI Press. 5.
- Hauser, S. (1976). "Loevinger's model and measure of ego development: A critical review." *Psychological Bulletin*. 83, 5, 928-955.
- Hauser, S. (1993). "Loevinger's model and measure of ego development: A critical review, pt.II" *Psychological Bulletin*. 4, 928-955.
- Hoppe, C. (1972). "Ego development and conformity behavior." St. Louis, Washington University, doctoral dissertation.
- Kegan, R. (1982). *The Evolving Self*. Cambridge, MA: Harvard University Press.
- Kegan, R. (1994). *In Over Our Heads: The demands of modern life*. Cambridge, MA: Harvard University Press.
- Kegan, R. & L. Lahey (1984), Adult leadership and adult development, In B. Kellerman ed. *Leadership*, Englewood Cliffs NJ: Prentice Hall, pp199-230.
- Kohlberg, L. (1963). "The development of children's orientations towards moral order: I. Sequence in the development of moral thought" *Vita Humana*. 6, 11-33).
- Kohlberg, L. (1964). "Development of moral character and moral ideology" in M. Hoffman and L. Hoffman eds. *Review of Child Development Research*," Vol 1. New York: Russell Sage. 383-431.
- Kohlberg, L. (1984). *Essays on Moral Development*, Vol. 2, *The Psychology of Moral Development*. San Francisco: Harper & Row.
- Loevinger, J. (1982). *Ego Development*. San Francisco: Jossey-Bass.
- Loevinger, J. (1985). "Revision of the Sentence Completion Test for ego development" *Journal Personality and Social Psychology*. 48, 420-427.
- Loevinger, J. ed. (1998). *Technical foundations for measuring ego development: The Washington University Sentence Completion Test*. Mahweh, NJ: Lawrence Earlbaum Associates.
- Loevinger, J. & Hy, T. (1996). *Measuring ego development* Second Edition. Mahweh, NJ: Lawrence Earlbaum Associates.
- Loevinger, J & Wessler, R. (1970) *Measuring ego development*, Vols. 1 and 2. San Francisco, Jossey-Bass.
- Lorr, M. & Manning, T. (1978). "Measurement of ego development by sentence completion and personality test" *Journal of Clinical Psychology*. 34, 354-360.

- Merron, K., Fisher, D., & Torbert, W. (1987). Meaning making and management action. *Group and Organizational Studies*. 12, 3, 274-286.
- Molloy, E. (1978). "Toward a new paradigm for the study of the person at work: An empirical extension of Loevinger's theory of ego development." Dublin, Ireland, University of Dublin, doctoral dissertation.
- Novy, D. & Francis, D. (1972). "Psychometric properties of the Washington University Sentence Completion Test" *Educational and Psychological Measurement*. 52, 1029-1039.
- Novy, D. et.al. (1994). "An investigation of the structural validity of Loevinger's model and measure of ego development" *Journal of Personality*. 62, 1, 87-118.
- Piaget, J. (1952/1937). *The Language and Thought of The Child*. London: Routledge & Kegan Paul.
- Redmore, C. (1976). "Susceptibility to faking of a sentence completion test of ego development" *Journal of Personality Assessment*. 40, 6, 607-616.
- Redmore, C. & Waldman, K. (1975). "Reliability of a sentence completion measure of ego development: *Journal of Personality Assessment*. 39, 3, 236-243.
- Rooke, D. & Torbert, W. (1998). Organizational transformation as a function of CEOs' developmental stage. *Organization Development Journal*. 16, 1, 11-28.
- Torbert, W. (1987). *Managing the Corporate Dream: Restructuring for Long-term Success*. Homewood, IL: Dow Jones-Irwin.
- Torbert, W. (1991). *The Power of Balance: Transforming self, society, and scientific inquiry*. Newbury Park, CA: Sage.
- Torbert, W. (1994). Cultivating post-formal adult development: Higher stages and contrasting interventions. In M. Miller and S. Cook-Greuter, eds. *Transcendence and Mature Thought in Adulthood: The Further Reaches of Adult Development*. Lanham MD: Rowman & Littlefield. 181-203.
- Torbert, W. & D. Fisher 1992. Autobiographical awareness as a catalyst for managerial and organizational development. ***Management Education and Development***, 23. 184-198.
- Vaillant, G. (1977). *Adaptation to Life*. Boston: Little Brown.
- Vaillant, G. & McCullough, L. (1987). "The Washington University Sentence Completion Test compared with other measures of adult ego development" *American Journal of Psychiatry*. 144, 9, 1189-1194.
- Westenberg, M., Jonckheer, P., Treffers, P. & Drewes, M. (1998). *Personality development: Theoretical, empirical, and clinical investigations of Loevinger's conception of ego development*. Mahwah NJ: Lawrence Erlbaum.