INVESTIGATING SELF ASSESSMENT USING DUAL SCALING

Dr. Betty McDonald

University of Trinidad and Tobago
O'Meara Campus, O'Meara Industrial Estate,
Arima, Trinidad, West Indies

Publication Date: August 15, 2011

Dr. Betty McDonald
University of Trinidad and Tobago, O'Meara Campus,
Lot 74-98 O'Meara Industrial Estate, Arima, TRINIDAD, WI.
Postal Address: 17 Portugal Crescent, Santa Rosa Heights,
Arima, TRINIDAD, WEST INDIES
Email: betmcdee@yahoo.com, betty.mcdonald@utt.edu.tt, betmcdee@gmail.com
1 (868) 642-8888, Ext. 21104 (Tel.), 1 (868) 643-1617 (Fax)
Running Head: Investigating Self Assessment using Dual Scaling.

INVESTIGATING SELF ASSESSMENT USING DUAL SCALING

ABSTRACT

This paper investigates self assessment using dual scaling which is a descriptive, optimal, non linear multidimensional and multivariate method that may be used to analyze nonlinear multivariate data 'very effectively and exhaustively'. Self assessment was defined as 'the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they have met these criteria and standards'. The participants for the study comprised 515 high school students (approximately aged 14-16 years old), representing more than a quarter (25.2%) of the school leaving cohort under study. Two hundred and thirty three were males. All participants were engaged in the same programme of studies. Dual scaling accounted for individual differences instead of averaging out responses and facilitated the examination and judgement of unexpected 'attitude' categories. The effect of individual differences in a study of this kind is important because fundamental variations among and between individuals are often obscured by averaging. Observations have implications for research where individual differences are important.

Keywords: Dual scaling; self assessment; individual differences; academic achievement.

INTRODUCTION

For the purpose of this paper, self assessment was defined as 'the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they have met these criteria and standards' *(*Boud, 1986:5). Boud posits two key elements as essential to every assessment (whether conducted by teacher or learner): (1) development of knowledge and an appreciation of appropriate standards and criteria for meeting those standards (2) capacity to make judgments about whether or not the work involved does or does not meet these standards (involves critical thinking).

Self assessment as operationally defined not only encompasses testing/grading one's own skills/work but involves an active process of evaluating what is good, mediocre or poor work in any given situation. 'Monitoring is the hub of self-regulated task engagement and the internal feedback it generates is critical in shaping the evolving pattern of a learner's engagement with a task' (Butler & Winne, 1995). Whilst it may not be unacceptable in some cases to assume that an aggregate score is representative of a participant's ability, for self assessment there is need to obtain as much information as possible about the individual if worthwhile judgements could be made for future use. This is particularly important when one considers the multifaceted developmental nature

of self assessment. The defining feature of self assessment is that the individual learner ultimately makes a judgment about what has been learned, not that others have no input to it (Boud, 1995). Self assessment may be viewed as the act of evaluating or monitoring one's own level of knowledge performance, and understanding in a metacognitive framework, taking into account the contexts in which it occurs. Self assessment involves the individual making an assessment of his or her own work, with an appreciation for and the understanding of those concepts of quality upheld and practised by the adjudicators of his or her work. The honing of self assessment skills would not naturally be endowed upon an individual but requires formal training, like any other skill (McDonald, 2010).

One is often interested in examining in detail the performance of a given participant on all items of an instrument. Because of the multifaceted nature of self assessment and the detection of individual differences this researcher thought it appropriate to use dual scaling for data analysis. This present paper therefore seeks to show how the technique of dual scaling may be used to obtain maximum information from a self assessment questionnaire.

The technique of dual scaling allows for the analysis of a wide variety of categorical data types like that from paired comparison, rank order, contingency tables and multiple-choice items. Dual scaling can capture not only linear relationships but quadratic, quartic, *etc*. Dual scaling assumes different names when applied to different data types. For instance when applied to a contingency table, dual scaling goes by the name ‘correspondence analysis’. When applied to multiple-choice data involving more than two items dual scaling is often referred to as ‘optimal scaling’ or ‘multiple correspondence analysis’ (Nishisato, 1994). In this present study we seek to use dual scaling to investigate self assessment.

LITERATURE REVIEW

Using our operational definition of self assessment stated earlier the self assessment formally taught by the classroom teacher initially represents the lowest level (receiving) of the affective domain. As time progresses and students internalize self assessment skills, higher levels of the affective domain would replace lower levels and students would embrace self assessment as a necessary and sufficient part of their daily activities. Developmental trends in self assessment have been reported by Van Krayenoord & Paris (1997). Despite the fact that children can start using self assessment to evaluate their achievements when quite young, older students are more effective at the process. There are differences within older students according to their levels of ability and the quality of teaching practices in particular classrooms. Metacognitive abilities associated with reading determine the quality of self assessments. Greater development in students' metacognitive abilities manifested itself in better ability for self-reflection and self-regulation of learning (Van Krayenoord & Paris, 1997).

Effectiveness of SA and self-management of learning improves with age, experience, intelligence, academic achievement and the quality of instruction (Paris & Cunningham, 1996; Van Krayenoord & Paris, 1997). Self assessment assists students to "learn how to learn". As students develop they rely less on the authority of grades and adults' evaluations as the sole source of feedback about their performance so that self assessment is foundational to the development of intrinsic motivation and autonomous learning.

In judging their own achievements, as children grow up they gradually change from equating achievement with 'effort' and see it related more to ‘ability' (Van Krayenoord and Paris, 1997; Blumenfeld, Pintrich, Meece and Wessels, 1982; Stipek and MacIver, 1989). This researcher has observed that by comparison with others, students of high ability tend to underestimate their own performances while students of lesser ability tend to overestimate their performances. However, when students focus their self assessments on clear criteria and standards this tendency was diminished (Van Krayenoord & Paris, 1997). Orsmond, Merry & Reiling (1997) confirmed that "good" students tended to underestimate their performance while "poor" students tended to overestimate it. Students producing good work were more self-critical than they were judgmental, whereas students producing poor work were less critical but more judgmental. Students handing in good assignments addressed both strengths and weaknesses of their work while others Self assessment is also a social activity occurring in situations that are social and collaborative and frequently with others who are more expert than the self assessor. Van Krayenoord & Paris (1997) noted that self assessment does not occur in isolation because the self has very little meaning unless it relates to others. This inevitably means that there must be a relationship between peers and teachers. The reliability and validity of self assessment is formulated not only in relation to criteria but also in relation to social interactions with assessments of peers and teachers. Hence, self assessment is an interactive, collaborative process involving the self and others in relation to criteria and standards. For this reason students who received formal training in self assessment skills in this present study were encouraged to discuss with their neighbors and arrive at mutually agreeable solutions to problems. Collaboration is the key to success. This explains why the researcher conducted group

sessions throughout the three terms of the academic year of self assessment training. Teachers and students became partners in the process of assessment and learning. The questions that teachers asked a class served as models for questions that learners asked themselves in self assessment. Educational goals underpin the questions and students are led, at different levels, to a realization of these goals (Van Krayenoord & Paris, 1997).

There are several problems related to human assessment (Richard and Helmes, 2000) and in particular self assessment. Because of its very nature there is variability amongst students themselves (catering for individual differences), particularly in terms of personality traits associated with self-esteem and self-concept. Some students have difficulty in commending themselves whilst others find it challenging to critically evaluate their performances (Bourke and Poskitt, 1997). In cases where self assessment relies on qualitative responses, the student's ability (or inability) with language may distort the intrinsic value of the assessment. Other issues are the truthfulness of self-reports. Hence, it is also problematic to arrive at consistency and so inevitably there are threats to the reliability of such assessments.

McAlpine (2000) extols the virtues of self assessment. He posits that it encourages metacognitive abilities and critical evaluation of the learner's educational goals while promoting student autonomy and decision making. He elaborates that self assessment also acknowledges choices and preferences in student learning styles and while being particularly relevant for open-ended learning activities it encourages intrinsic motivation and self-sustained learning. McAlpine (2000) sees self assessment as encouraging success and lifelong learning, developing students' responsibility for their own learning and encouraging a collaborative student-teacher relationship in learning and assessment. He further elaborates that self assessment may also be used to

determine existing competencies and it is a useful and individualized way of enabling students to establish if their prior learning is relevant for their next learning activity. Consequently, students, especially those with special abilities could avoid wasting time studying material they have already covered. Finally, self assessment is also relevant as a means of evaluating whether major learning goals have been met in learning contracts that are often used in the education of students with special abilities (McAlpine, 2000).

In support of the unreliability of self assessment Dunning *et al* (2004) affirm that the act of self-assessment is an intrinsically difficult task. Their reviews from psychological inquiry research found that skill and character self assessments demonstrate both substantive and systematic flaws. In particular they focused on the implications of such flawed self assessments in three real-world domains, viz. health, education and the workplace. Dunning *et al* (2004) further enumerated several obstacles that prevent people from revealing truthful self-impressions. They affirm that in general several psychological processes conspire to produce flawed self-assessments. People claim to be what they are not and often behave in ways contrary to what they purport. Over estimations are prevalent, self promises are often not kept and outcomes are more optimistic than realistic. Nevertheless they encourage both researchers and practitioners to recognize the multi disciplinary nature of self assessment and its significance in human and institutional development. In the same vein they caution policymakers and other people who make real-world assessments to be wary of self-assessments of skill, expertise and knowledge. To avert the full effect of human error stakeholders should consider ways of 'repairing' flawed self assessments.

It seems reasonable that the multi-faceted nature of self assessment (embodying numerous disciplines like psychology, sociology, education, engineering, medicine,

accounting, business, *etc*.) as briefly described earlier demands a technique that is able to obtain as much information as possible from the data and highlight individual differences at the same time. Current literature offers many traditional techniques for assessing the effectiveness of an intervention. However, dual scaling appears to fill a gap that allows for the exposure of the multifaceted dimensions of a construct like self assessment. This researcher believes that there are as many (or perhaps more) differences between people as there are similarities. The effect of individual differences in a study of this kind is important because fundamental variations among and between individuals are often obscured by averaging. For example, a researcher interested in SA in students may use a sample of students, measure their self assessment skills and arrive at a single average, presumably indicative of the entire group. The researcher then disseminates information about the entire population without due regard to individual differences that would take account for age, sex, ethnic background, ability, academic aptitude, personality, motivation, self efficacy, self esteem, interests and other factors that influence self assessment. Such an average reported based on the results would shroud the multidimensional nature of self assessment and mislead readers. Most psychological research depends upon statistical analyses that are applicable to groups of people. However, when there is an interest not only in characteristics shared by all individuals of a group but also specific individuals a technique that accounts for individual differences is mandatory.

Hence, this present study investigates self assessment using a technique that addresses individual differences. To summarise very briefly, dual scaling is a descriptive, optimal, non linear multidimensional and multivariate method or technique that may be used to analyze nonlinear multivariate data '*very effectively and exhaustively*'. It extracts

quantitative information from non-numerical (qualitative) data. Dual scaling assigns simultaneous weights to response options and scores to participants in such a way as to optimize Guttman's principle of internal consistency. The process for dual scaling operates in the following manner:

> 'Assign as similar scores as possible to those subjects who chose the same option of a question, and these scores should be as different as possible from the scores of those who chose other options. Assign as similar weights as possible to those options which are chosen by one subject and these option weights should be as different as possible from the weights of options which are not chosen by this subject' *(*Nishisato, 1994:3*)*.

Because the sum of responses weighted by row is equal to the sum of responses weighted by column, which is set at zero, it means that the sum of the weighted responses of each item is zero. With no mean differences between the items, one can readily distinguish item difficulty or popularity. Option weights are determined to maximize the average of the sums of squares of item-total correlations or to maximize the internal consistency reliability alpha. Maraun, Slaney & Jalava (2005) successfully used dual scaling to explain in nontechnical terms what it offered to an analysis using contingency table and multiple-choice data.

THE STUDY

The objective of the present study was to use the technique of dual scaling to unravel the multifaceted construct self assessment. The study essentially entailed the administration and analysis of responses from a 64-item questionnaire, patterned after Jackson Personality Research Form (PRF)-Form E. A total of 16 items were taken from each of the four scales (achievement, autonomy, endurance and understanding). The achievement and endurance scales measured orientation toward work, the autonomy scale

measured orientation toward direction from other people and the understanding scale measured intellectual orientation.

METHOD

Participants and context

The participants for the study comprised 515 high school students (approximately aged 14-16 years old), representing more than a quarter (25.2%) of the school leaving cohort under study. Two hundred and thirty three were males. All participants were engaged in the same programme of studies. The participants were chosen from a Caribbean island which is one of the 16 English speaking Caribbean territories that participate in an external regional examination that is compulsory for all high school students. Serving 16 territories and offering examinations in about 62 different subject areas, the external regional examination board is the only one in the Caribbean region.

Selection process

Participants were randomly sampled from ten high schools representing top, middle and bottom levels of achievement, rated according to their performance at the criterion referenced external examination results for three consecutive years; parental choice of high school based on the results of the secondary entrance examination that qualifies entrants for entry into those high schools and official comments from educational officers at the ministry of education.

Instrument

Participants were required to respond to 64 statements in a pen-and-pencil survey. A copy of the questionnaire (SA2) is available from the author upon request. Basically the questionnaire solicited dichotomous responses (yes/no) to varied statements like 'There are more activities I prefer to reading'; 'Family obligations make me feel important'; 'I often set goals that are very difficult to reach'; 'I like to do whatever is proper'; 'Nothing would hurt me more than a bad reputation'; "I try to work just hard enough to get by'; 'Studying the history of ideas has no appeal to me' and 'I do not let my work get in the way of what I really want to do'. These statements were previously classified as belonging to one of the four Jackson's scales (achievement, autonomy, endurance and understanding). There were no right and wrong responses as items were scored as yes' or no's. The original data from the results of the self assessment instrument (SA2) was coded in dual scaling format (series of zeros for no's and ones for yes'). This specially coded format of zeros and ones was given the name SA2DUAL3 to distinguish it from any other data set at hand. The computer software program DUAL3 (Nishisato, 1994) was used for data analysis.

RESULTS

Using the data gathered from the sample SA2DUAL3 showed conformity to a two factor model identified for the construct self assessment. Bear in mind that the 64 item instrument was tested on a variety of different groups other than this specific cohort. It was therefore not surprising that a two factor model emerged. Results indicated reasonable reliability or internal consistency (.61) for the first factor designated an anti intellectual disposition (AID) factor, accounting for a 40 % of the variance and reliability

of .54 for the second factor designated social and family commitment (SFC), accounting for 33% of the variance (Table 1).

<Table 1 to be inserted here>.

Together the two factors accounted for 73% of the variance. Two other factors that were extracted for further confirmation of the two factor model for self assessment produced reliabilities that were unacceptably low (.32, .26 resp.) so threatening any further analysis with them. This meant that by using more than two factors the extent to which the response patterns of participants could be predicted from their scores and item difficulty scores was unreliable.

The correlation indicated the amount of linear relationship between items of which options are optimally scaled. Optimum weights are determined in order to maximize the sum of squares of all the inter-item correlations. The alpha (Table 1) is really the reliability coefficient alpha or the generalized Kuder-Richardson reliability or internal consistency reliability or Cronbach's generalizability coefficient alpha. It gives the extent to which "high scorers" choose options with large weights and "low scorers" choose options with low weights (Nishisato, 1994: 44). While the alpha coefficients may appear low (indicative of a high error of measurement) it must be remembered that SA, by its very multifaceted nature demands multi forms of assessment. In this experiment a pen-and-paper instrument in the form of a questionnaire was used to provide the data for analysis. Delta (Table 1) indicates the total variance explained by the solution. The percentage homogeneity (Table 1) indicates the degree to which the derived solution conforms to the case of perfect consistency. It is also an index of how good the solution

is as compared with the perfect case where all the inter-item correlations are one or a single item containing all of the information in the data.

The variance/covariance matrix (available upon request from the author) supported the identity of the factors. The residuals of the order of 0.001 after fitting a two dimensional model indicated how well the model fitted and showed how the principle of local independence was satisfied. Unexplained causes, for example, random variation in scores and systematic components for which no suitable predictors are provided were sighted as possible sources of error.

Dual scaling provided details about the individual items that may be used to reconstruct an improved self assessment instrument. The contribution of each item to explaining the data is reflected in the relevance of the item to the data (r_{jt}). For example, in the self assessment instrument –SA2 (available upon request from the author), items 3 ('I enjoy difficult work') ; 19 ('I often set goals that are very difficult to reach'); 36 ('As a child I worked a long time for some of the things I earned'); 41 ('I think I would enjoy studying most of my life so I could learn as many things as possible') and 42 ('My goal is to do at least a little bit more than anyone else has done before) are good items for solution 1 whilst items 1 ('"There are many activities I prefer to reading'); 5 ('I don't have the staying power to do work that must be very accurate'); 6 ('Family obligations make me feel important') are among the poor items. For a poor item it does not make much of a difference which option of the item is chosen, while for a good item one can clearly distinguish between two participants (individual differences) who choose different options. This information may be used to design a more effective self assessment instrument. This meant that the extent to which the response patterns of participants can be predicted from the scores and item difficulty scores was reliable. It must be noted that

dual scaling determines option weights so as to maximise the internal consistency reliability alpha.

CONCLUSION

This paper addresses an issue of potential interest in educational psychology. Stepping outside traditional classical test theory methods to dual scaling gives it a unique flavour that is of significance to researchers. The items of the 64-item self assessment questionnaire SA2 were patterned after The Jackson Personality Research Form (PRF)-Form E that had been successfully tested over an extended period of time on a variety of different population groups, for example, male students, female students, juvenile offenders and adults, psychiatric patients, college students and personnel from the military. With no available benchmarks for measuring self assessment, the panel felt that it was best to use this universally accepted instrument as an approximate gauge for the measurement of the construct self assessment after formal training in self assessment skills. Additionally, the universality of the Jackson Personality Research Form (PRF)-Form E would offer better external validity especially because readers may wish to generalize the results to their own contexts.

Whilst Jackson did not construct the scales as criteria but as measures for norm-oriented measurement of certain needs as personality traits, this researcher perceived the scales as a close enough representation of the essential attributes of self assessment as defined by Boud (1986). The point must be clearly made here that self assessment involves action. The panel of measurement experts agreed to use trait-like scales from a

personality inventory to gauge or quantify or evaluate the key attributes of the construct self assessment. For example, the scale achievement in the context of self assessment refers to whether or not participants are able to identify appropriate standards. Defining trait adjectives for achievement in the Jackson's PRF are 'striving, accomplishing, capable, purposeful, attaining, industrious, achieving, aspiring, enterprising, self-improving, productive, driving, ambitious, resourceful, competitive'. Whilst the intention to meet standards and qualities like aspiring to accomplish difficult tasks, maintaining high standards and willing to work toward distant goals may not be absolutely equivalent, the panel of measurement experts agreed that generally (given approximately equivalent mental abilities and controlling for other variables like preknowledge, attitude, *etc*.) participants with the aforementioned qualities are able to identify standards better than those participants who use less of those said qualities.

Similarly, autonomy in the Jackson's PRF indicates an 'unmanageable, free, self-reliant, independent, autonomous, rebellious, unconstrained, individualistic, ungovernable, self-determined, non-conforming, uncompliant, undominated, resistant, lone-wolf person'. The present study assumes that these characteristics would enable a person to better identify appropriate standards and/or criteria and apply those standards and/or criteria to making judgements about work done. Recognizing that a 'manageable, not free, not self-reliant, dependent, nonautonomous, not rebellious, constrained, nonindividualistic, governable, not self-determined, conforming, compliant, dominated, nonresistant, gregarious' person could well have the capacity to identify appropriate standards and/or criteria and apply those standards and/or criteria to making judgements about work done, the panel of measurement experts took the position that the comparison

of capacities is under scrutiny rather than the absolute capacities. The degree to which this enablement is possible is considered as a necessary limitation of the present study.

A similar discussion applies to the scales understanding and endurance. It is also recognized that whilst the attributes for understanding and endurance may be necessary conditions for successful self assessment they need not be sufficient conditions. Taken together, whilst there may be no absolute need for a person to correctly self assess to be a high achiever, to be autonomous, to be enduring or to be understandful, the panel of measurement experts agreed that there is a higher probability for a person who is so disposed to identify standards and/or criteria than there is for a person who is not or for a person who is so disposed to a lesser degree. In the present study the panel of measurement experts agreed that the interrelatedness of the selected scales together described self assessment as defined by Boud (1986). The non linearity of the scales is also recognized and treated as an assumption and a necessary limitation. High scores on a given scale are assumed to be better indicators of ability to identify appropriate standards and/or criteria and apply them to making judgements about work done whilst this may not necessarily be so in the purest sense. Leaving aside the issues related to the alignment of scales, the present study focuses on two core issues of a possible measurement instrument, namely, how to (1) set standards and /or criteria to monitor value-added performance and productivity of participants across subject disciplines and (2) use those set standards and/or criteria to make judgements about work done.

DISCUSSION

Some mention must be made here of the issue of aggregation. Aggregation has come to be known as the process of combining observations of educational performance

into a single indicator to serve two main technical functions. One function is to enhance the reliability of the indicator because random errors in the individual observations tend to cancel each other out during the aggregation process. The other function is to enhance the validity of the indicator by enabling the assessment process as a whole to sample the domain of achievement being assessed. Oftentimes, conventional aggregation processes allow differing weights to be given to different areas of the domain in order to define the nature of the attainment summarized by the indicator.

No weights have been applied to the different statements since each statement is taken to be of approximately equal importance in the definition of self assessment. Participants having similar scores on any of the four scales may have responded to yes or no, in different ways, within that same scale. Even participants having overall parallel profiles may have different scores across scales. Within-scale variability, seen as a limitation to the generalized results of the procedure was addressed by use of non linear factor analysis using the computer software package DUAL3. For the SA2 instrument, all statements were selected directly from the four predetermined scales, *viz.* achievement (Ac), autonomy (Au), endurance (En) and understanding (Un) from The Jackson Personality Research Form (PRF) - Form E. The latter was drawn from extensive review of the literature, comments from students, comments from teachers, comments from research specialists and a pilot study of the population.

In the Jackson's Personality Research Profile (PRF) - Form E, item selection with a concern for the avoidance of a preponderance of items of psychopathological content (common to many such scales), assumed preeminence. For example, from a large pool of items that had previously been screened for desirability, approximately 16 from each extreme of the distribution of desirability scale values were chosen. Those items that

showed a significant amount of content similarity or homogeneity were eliminated. For general unambiguity and absolute clarity and intelligibility, the chosen items were rewritten. Using a total of 107 items administered to 305 male and female secondary (high) school students, it was found that these items were substantially consistent (Kuder-Richardson formula 20 reliability estimate = .83) with the responses of those previously scaled items.

With official permission obtained for use, the Jackson's Personality Research Profile (PRF) - Form E instrument selected was considered ideal because it was designed to yield scores for personality traits relevant to the normal functioning of individuals rather than psychopathological behaviour. The participants for this present study, by and large are normal functioning individuals who have demonstrated their potential for further study by their previous academic achievement.

The original set of variables defined by Murray, 1983 (cited in Jackson, 1984) and his colleagues at the Harvard Psychological Clinic for comprehensively describing personality and those found in the literature formed the nucleus of item generation for The Jackson's Personality Research Profile (PRF) - Form E instrument.

Using the Delphi Technique, measurement and evaluation specialists considered the four scales achievement (Ac), autonomy (Au), endurance (En) and understanding (Un) relevant to self assessment. To proceed with the four scales selected interrator reliability was determined using three approaches (1) percentage agreement (Kappa .97) (2) consistency correlations (Pearson's r) and (3) decision consistency (Buckendahl *et al* (2003). Bearing in mind that operationally self assessment was defined as 'the involvement of students in identifying standards and/or criteria to apply to their work

and making judgements about the extent to which they met these criteria and standards' (Boud, 1986: 5), all of the 22 scales from the Jackson Personality Research Form (PRF) -Form E were scrutinized to ensure that they attempted to measure the extent to which students could identify standards and/or criteria to apply to their work. Further, the measurement and evaluation specialists had to carefully choose the scales in order to ensure that as much as possible students would be able to develop skills in deciding the extent to which such judgements met the criteria previously chosen (definition of self assessment). For example, in the choice of the Au scale, consideration was given to the fact that to be able to identify criteria or standards that apply to their work, students had to be willing to break away from restraints, confinements or restrictions and enjoy being able to act independently without the intervention of a teacher. Self-reliance, self-determination, independence and individualism are needed to make judgements within the context of the previous criteria. Clearly, self assessment promotes synergy among the different constructs to create a holistic discipline that is greater than the sum of its components. The measurement and evaluation specialists used similar reasons for their choice of the other scales. The items on each scale may appear to measure other constructs like self-efficacy, self-esteem, self-concept, *etc*. but one has to appreciate that the construct self assessment encompasses a number of qualities that affect the whole individual (Gestalt view). Self-efficacy, self-esteem, self-concept, *etc*. are important in helping an individual identify standards or criteria to apply to his or her work and make judgements about the extent to which he or she has met these criteria and standards (self assessment definition).

It is clear that self assessment is not only multifaceted but individualized and plays an integral part in the learning process and student achievement. As individuals

students bring to the table a host of personal concerns having come from different backgrounds and cultures and having being brought up under unique circumstances. This particular study simplified self assessment by focusing on somewhat 'linear', non overlapping dimensions that do not exist in a unilinear manner in real life situations. There is a tremendous amount of overlapping of dimensions that are difficult to unravel in a manner that would make sense of authentic information. Accordingly, any investigation of self assessment requires a technique like dual scaling that highlighted the detailed information that can be obtained from data especially where individual differences are required to be captured. Because self assessment is multifaceted and encompasses several features as described earlier and focuses on the participant as an individual, the minutest differences among participants need to be captured for meaningful interpretation of the concept. The results of the present study demonstrate how dual scaling may be used to maximize the amount of information obtainable particularly where detailed individual differences are important. It is hoped that more use would be made of dual scaling in further research aimed at understanding the role of individual differences within an educational context.

BIBLIOGRAPHY

Adams, C. & King, K. (1995). Towards a framework for student self-assessment. *Innovations in Education and Training International, 32*, 4, 336-343.

Blumenfeld, P.C., Pintrich, P., Meece, J. & Wessels, K. (1982). The formation and role of self perceptions of ability in elementary classrooms. *Elementary School Journal 82*, 401-420.

Boud, D. (1986). *Implementing Student Self Assessment*, Sydney: HERDSA.

Boud, D. (1989). The role of self assessment student grading. *Assessment and Evaluation in Higher Education, 14*, 1, 20 –30.

Boud, D. and Falchikov, N. (1989). Quantitative studies of student self assessment higher education: a critical analysis of findings. *Higher Education, 18*, 529 –549.

Boud, D. (1995). *Enhancing Learning through Self Assessment*. London: Kogan Page Limited.

Bourke, R. & Poskitt, J. (1997). *Self Assessment in the New Zealand Classroom* (Booklet). Wellington: Ministry of Education.

Buckendahl, C. W., Yongwei, Y. & Abdullah, F. (2003). 'An Alternative Strategy for Estimating Decision Consistency Reliability". Paper presented at the annual meeting of the American Educational Research Association (AERA)/National Council on Measurement in Education (NCME), April 21-25, 2003, Chicago, Illinois, USA.

Buss, D.M. & Greiling, H. (1999). Adaptive Individual Differences. *Journal of Personality, 67*, 209-243.

Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis, *Review of Educational Research, 65*, 3, 245-281.

Dunning, D. , Heath, C. and Suls, J.M. (2004). Flawed Self-Assessment Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest.* 5: 69-106,doi:10.1111/j.1529-1006.2004.00018.x

Eysenck, H. J. & Eysenck, M.W. (1985). Personality and Individual Differences: A natural science approach. New York: Plenum.

Jackson, D. N. (1984). *Personality Research Form Manual*. Michigan: Research Psychologists Press Inc.

Jackson, D. N. (2002). "Personality Research Form". Sigma Assessment Systems. <http://www.sigmaassessmentsystems.com/prf..htm> (March 05, 2010).

Lucas, R. & Baird, B. (2006). Global self assessment. In M. Eid & E. Diener (Eds.), Handbook of psychological measurement: A multimethod perspective (pp. 29 -42). Washington, DC: American Psychological Association.

Maraun, M. D., Slaney, K. & Jalava, J. (2005). Dual Scaling for the Analysis of Categorical Data. *Journal of Personality Assessment, 85*, 2, 209-217.

McAlpine, D. (2000). Assessment and the Gifted. *Tall Poppies, 25 (1).* http://www.tki.org.nz/r/gifted/pedagogy/tallpoppies_e.php. (March 03, 2010).

McDonald, B. (2010). Self Assessment and Academic Achievement. ISBN-NR: 978-3-8383-3360-1 :Lambert Academic Pub AG & Co. KG. Germany: Saarbrucken.

Nishisato, S. (1988). *Dual scaling of Multiple-Choice Data (DUAL# for Windows, Version 4.10=(32bit)).*

Nishisato, S. & Nishisato, I. (1994). *Dual scaling in a Nutshell.* Toronto:Microstats.

Orsmond, P., Merry, S. & Reiling, K. (1997). "The use of student derived marking criteria in peer and self assessment". *Assessment and Evaluation in Higher Education, 25* (1), 23-38.

Paris, S.G. & Cunningham, A. (1996). Children becoming students. In D. Berliner and R. Calfee (Eds.), *Handbook of Educational Psychology*. New York: Macmillan, 17-147.

Pike, G. (2006): The Convergent and Discriminant Validity of NSSE Scalelet Scores. *Journal of College Student Development, 47*, 551-564.

Richard, G. D. & Helmes, E. (eds.) (2000). *Problems and solutions in human assessment: honoring Douglas N. Jackson at seventy*. Boston: Kluwer Academic Publications.

Schweizer, K. (2005). An Overview of Research into the Cognitive Basis of Intelligence. *Journal of Individual Differences, 26*, 1, 43-51.

Snow, R. E., Corno, L. & Jackson, D. (1996). Individual Differences in Affective and Conative Functions. In D.C. Berliner and R.C. Calfree (Eds.), *Handbook of Educational Psychology*, 243-310 (Part 1 and 2).

Stipek, D.J. & MacIver, D. (1989). Developmental changes in children's assessment of intellectual competence. *Child Development, 60*, 521-538.

Van Krayenoord, C.E. & Paris, S.G. (1997). Australian students' self-appraisal of their work samples and academic progress. *Elementary School Journal 97,* 5, 523-537.

Table 1

Summary statistics for SA 2 using dual scaling from the computer software package DUAL3

Solution	Factor 1 (AID)	Factor 2 (SFC)
Correlation Ratio	0.04	0.03
Maximum Correlation	0.20	0.18
Alpha	0.61	0.54
Delta	7.23	6.14
Cumulative Delta	7.23	13.37
Delta B	0.30	0.25
Cumulative Delta	0.30	0.56
% Homogeneity	3.90	3.31

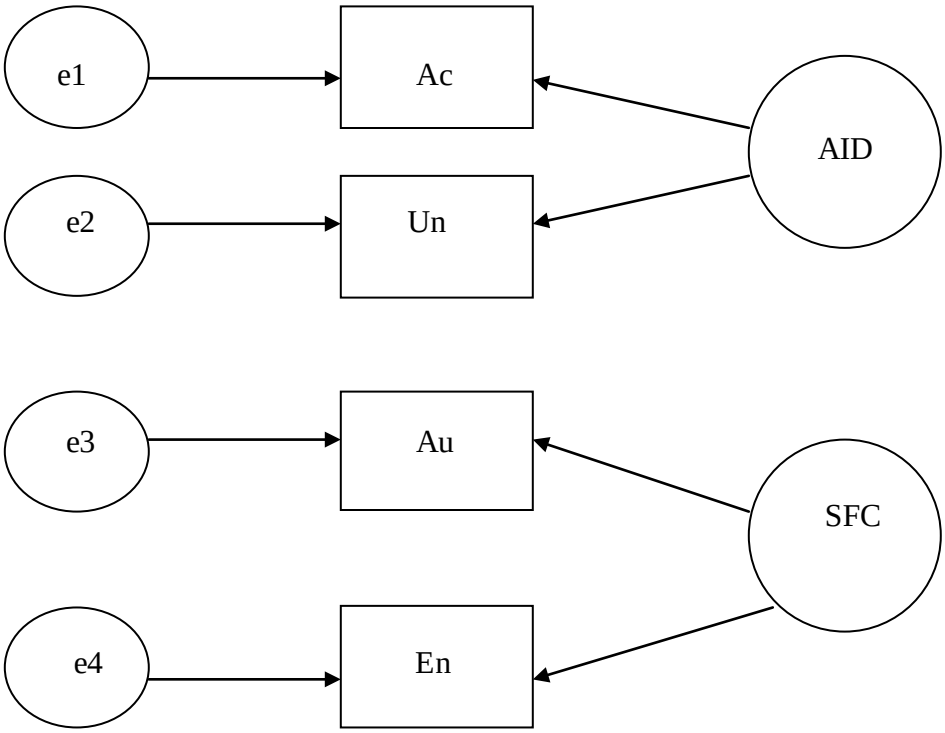

Figure 1 Proposed model for self assessment