Australian Council for Educational Research (ACER)

ACEReSearch

2004 - The International Test Users' Conference   1997-2008 ACER Research Conference Archive

2004

International Test Users' Conference: Assessing intelligence, emotion and behaviour (Conference Proceedings)

Australian Council for Educational Research (ACER)

Follow this and additional works at: https://research.acer.edu.au/research_conferenceITU_2004

Part of the Educational Administration and Supervision Commons

Recommended Citation

Conference Proceedings

Contents

Foreword

Geoff Masters

Australian Council for Educational Research

Professor Geoff Masters is Chief Executive Officer of the Australian Council for Educational Research (ACER). Professor Masters is an international authority in educational measurement and student assessment and has published extensively in these fields. Early in his career he developed the widely-used partial credit model for the statistical analysis of rating scales and professional judgements. Although much of his research has been focused on questions of validity and reliability in large-scale tests and surveys, Professor Masters has a special interest in using developments in modern measurement theory to construct improved tools for professional practitioners.

The Australian Council for Educational Research (ACER) is delighted to host the 2004 International Test Users Conference.

This year's conference, which includes participants from the United States, Malaysia, Indonesia and New Zealand, brings together an impressive group of world leaders in the assessment of intelligence, behaviour and emotions. Keynote presentations will set the scene for a comprehensive understanding of people, their capacities and capabilities, and methods for their assessment.

Conference presenters will address issues in assessing the whole person across the lifespan. They include:

- Professor Gale Roid, co-author of the Stanford Binet 5, a world-renowned assessment tool which provides a single solution to intelligence testing needs for ages 2 to 85+ years;

- Professor Thomas Achenbach, author of the Child Behaviour Checklist, an assessment tool that provides an integrated multi informant (self, teacher, parent) assessment of behaviour aligned to the Diagnostic and Statistical Manual of Mental Disorders fourth edition;

- Dr David Caruso, co-author of the Mayar, Salovey, Caruso Emotional Intelligence Test, the first ability-based measure of emotional intelligence which assesses capacity to reason using emotional information; and

- Dr Katherine Hirsh, co-author of a number of texts on the Myers Briggs Type Indicator, an influential tool with 60 years of theory, research and practice that provides contemporary solutions to organisational and relational issues.

Keynote addresses will be well supported by a range of concurrent sessions on leadership, executive coaching, online assessment, organisational behaviour, the use of psycho-behavioural testing, and applications of intelligence, behaviour and emotional intelligence testing.

This year's International Test Users Conference with its depth of focus – from adaptive to maladaptive behaviours; from gifted children to low functioning non-verbal children – will be equally relevant to education professionals, organisational and human resource professionals, psychologists and those interested in personal discovery.

Geoff N Masters

Professor Geoff N Masters
Chief Executive Officer, ACER

Plenary Papers

Psychopathology and Adaptive Functioning Across the Life Span: *Top-down, bottom-up, multi-informant, and multi-cultural challenges and solutions*

Thomas M. Achenbach

University of Vermont

Thomas M. Achenbach, Professor of Psychiatry and Psychology, is Director of the Center for Children, Youth, and Families at the University of Vermont Department of Psychiatry. A summa cum laude graduate of Yale, he received his Ph.D. from the University of Minnesota and was a Post-doctoral Fellow at the Yale Child Study Center. Before moving to the University of Vermont, Dr. Achenbach taught at Yale and was a Research Psychologist at the National Institute of Mental Health. He has been a DAAD Fellow at the University of Heidelberg, Germany; an SSRC Senior Faculty Fellow at Jean Piaget's Centre d'Epistémologie Génétique in Geneva; Chair of the American Psychological Association's Task Force on Classification of Children's Behavior; and a member of the American Psychiatric Association's Advisory Committee on DSM-III-R. He has given over 200 professional presentations and has authored over 200 publications, including Developmental Psychopathology; Research in Developmental Psychology: Concepts, Strategies, Methods; Assessment and Taxonomy of Child and Adolescent Psychopathology; Empirically Based Taxonomy; Empirically Based Assessment of Child and Adolescent Psychopathology (with Stephanie H. McConaughy); and Manuals for the Child Behavior Checklist, Teacher's Report Form, Youth Self-Report, and other standardized assessment instruments. Dr. Achenbach's honours include the Distinguished Contribution Award of the American Psychological Association's Section on Clinical Child Psychology and the University Scholar Award of the University of Vermont.

This keynote address will outline some important challenges for assessment of psychopathology and adaptive functioning from age 1 to 90+ years. It will then present practical ways to meet the challenges.

Top-down and bottom-up challenges

One group of challenges concerns *top-down* and *bottom-up* models for psychopathology, differences and points of contacts between these models, how they engender taxonomies of problems, and their implications for service, research, training, and communication.

The *top-down* approach to psychopathology is exemplified by the World Health Organization's *International Classification of Diseases* (ICD) and the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM). In the top-down approach, committees of experts negotiate diagnostic categories to be included in a nosological system. After choosing the categories, the experts negotiate definitions of diagnoses and of criteria for making diagnoses. This approach is 'top-down' in the sense that it begins 'at the top' with experts' concepts of disorders and then works down to criteria for determining who has each disorder. The disorders are defined categorically and are judged to be either present or absent. People who meet criteria for a particular disorder are diagnosed as having that disorder. Moreover, people who meet criteria for multiple disorders are diagnosed as having each of the disorders for which they meet criteria. Findings that many people meet criteria for multiple disorders have generated a large literature on the 'comorbidity' (co-occurrence) of disorders.

In contrast to the top-down approach, the *bottom-up* approach starts with large pools of items for assessing problems. These items are used to assess large samples of people. Scores obtained on the problem items are subjected to multivariate statistical analyses in order to identify sets of problems that tend to co-occur. The sets of co-occurring problems are designated as *syndromes*. Each syndrome comprises problems that have been found to co-occur in large samples of people. The problems comprising a syndrome are analogous to the symptoms that are specified as criteria for diagnostic categories in top-down systems. However, in contrast to the present-versus-absent diagnostic model, syndromes are scored quantitatively to measure the *degree to which* each person manifests a particular set of problems. Thus, people obtain scores that vary on a continuum from low to medium to high. Cutpoints can be applied to the continuum of scores to mark particular ranges of scores, such as the normal, borderline, and clinical ranges. However, scores can vary within each range, as well as between ranges. Thus, people whose scores are in the normal range can differ with respect to whether they are in the low, medium, or high normal range. Similarly, people whose scores are in the clinical range can differ with respect to whether they are in the low, medium, or high clinical range. Furthermore, because people obtain scores on all syndromes that have been

found for their gender and age, they can have high scores on multiple syndromes without being assumed to have multiple disorders.

Profiles of syndrome scores graphically display the areas in which a person has mild, moderate, or severe levels of problems. Because many profile patterns are possible, the profile patterns can provide individualised pictures of people's problems. The fact that many people manifest a variety of problems is handled by documenting their particular patterns, rather than by equating deviance in multiple areas with comorbidity among different disorders. Profiles of scales for positive adaptive functioning also provide individualised pictures of people's strengths.

Current versions of the top-down and bottom-up approaches have several important points of contact. For example, they both provide fairly explicit descriptions of behavioural, emotional, thought, and social problems, rather than depending on inferences about characteristics that are unconscious or are impossible to observe for other reasons. In addition, some of the syndromes derived through bottom-up methodology comprise sets of problems like those used as criteria for certain top-down diagnostic categories. Numerous studies have reported statistically significant associations between top-down diagnoses and scores on bottom-up syndromes. Nevertheless, mismatches between top-down diagnostic categories and bottom-up syndromes and between the different assessment procedures employed by the two approaches present challenges for practitioners and researchers who wish to capitalise on the potential benefits of both approaches.

Multi-informant challenges

A second type of challenge concerns the sources of data for assessment of psychopathology and adaptive functioning. Although it is generally accepted that assessment of children requires information from parents as well as from the children themselves, correlations between reports by mothers and fathers are not high enough to ensure that a mother will typically provide the same picture of the child as the father will. Furthermore, correlations of parents' reports with reports by teachers, clinicians, and observers are far lower than correlations between mothers and fathers. Correlations between reports by the various non-parental informants are similarly modest, while correlations of children's self-reports with adults' reports of the children's functioning are still lower.

The low cross-informant correlations do not mean that any of the informants are inherently unreliable, as good reliabilities have been obtained for reports by each type of informant. Instead, the cross-informant correlations are likely to be limited by the different samples of children's behaviour observed by each informant. The correlations may also be limited by the informants' different effects on children, by their different personal perspectives, differences in their recall of the children's behaviour, and different degrees of candour.

It is clear that assessment of children requires data from multiple adults, as well as from the children themselves. However, assessment of adults typically relies on data obtained only from the adult client via interviews, questionnaires, and tests. In contrast to the many studies of cross-informant data on child psychopathology and adaptive functioning, relatively few studies have reported cross-informant correlations for assessment of adult psychopathology. To conduct meta-analyses of agreement between informants' reports on adult

psychopathology, we manually searched some 47,000 articles published in 46 peer-reviewed journals between 1993 and 2003. Of the 47,000 articles, only 103 (0.2%) reported cross-informant correlations that met minimal scientific standards. Meta-analyses showed that correlations between clients' self-reports and reports by people who knew the client averaged only about .40 for most kinds of problems. Furthermore, correlations between pairs of informants who knew the adult clients averaged only in the .20s. Thus, the challenges of obtaining and integrating data from multiple informants pertain to assessment of adults as well as children.

Although informants' reports are sometimes sought for assessment of elderly adults whose competence is in doubt, this is seldom done in a way that facilitates systematic comparisons between reports by elderly clients versus reports by informants who know them. However, using parallel self-report and informant-report instruments that will be described in the keynote address, we obtained correlations between self-reports and informant-reports that averaged .51 for people aged 60 to 102 years. As this mean correlation was actually higher than the mean cross-informant correlation of .40 obtained for 18- to 59-year-olds, it suggests that elders' self-reports can be very useful when systematically compared with informants' reports.

Multi-cultural challenges

For a host of reasons, assessment of psychopathology and adaptive functioning faces challenges with respect to cultural differences. Many countries need to provide educational, mental health, and social services for refugees and immigrants, who number

in the millions worldwide. Differences in language, culture, socioeconomic status, education, values, and expectations challenge traditional assessment practices. Increasing sensitivity to the cultures and needs of native-born minority groups also poses challenges for assessment of psychopathology and adaptive functioning.

It is unrealistic to assume that assessment methods developed in one culture are equally applicable to people of all cultures. It is also unrealistic to assume that every cultural group will develop its own culture-specific assessment methods. This is especially unrealistic in view of the blending of cultures that is occurring throughout the world. Furthermore, assessment methods tailored to a particular cultural group may quickly become irrelevant when members of that group move or are exposed to other cultures.

In addition to being sensitive to the cultures of people who are assessed, methods must also yield data that help the users make better decisions. Considering the variety of cultural groups that may be present in many countries, assessment professionals cannot be experts in the cultures of all those who they are called on to evaluate. Instead, solutions to multi-cultural challenges require methods that can be used by professionals to make decisions about members of different cultural groups who are served by the educational, mental health, and other services that employ the professionals.

Practical ways to meet the multiple challenges

With the collaboration of colleagues from many cultures over the past few decades, our research team has developed a family of assessment instruments that are designed to meet the various challenges. An overview of the instruments and solutions will be presented in this address, while details of the specific instruments and how to use them will be presented in the workshop by Drs. Achenbach and Rescorla.

In brief, the instruments include standardized forms for obtaining assessment data from multiple informants in developmentally appropriate ways for ages 1 to 5 years, 6 to 18 years, 18 to 59 years, and 60 to 90+ years. For children and adolescents, the instruments include versions designed for completion by parents, teachers, daycare providers, clinical interviewers, adolescents, psychological examiners who administer ability and achievement tests, and observers who record behaviour in group settings such as classrooms. For ages 18 to 90+, the instruments include self-report versions and informant versions, which can be completed by spouses, partners, relatives, grown children, friends, roommates, mental health workers, and others who know the person being assessed.

Top-down and bottom-up assessment

The assessment data yield scores for top-down DSM-oriented scales that were constructed by having mental health professionals from 20 cultures identify problem items that they judged to be very consistent with DSM diagnostic categories. Each DSM-oriented scale is scored quantitatively by summing the scores of the items that comprise the scale. The DSM-oriented scales are displayed on profiles in relation to age-specific norms for each gender.

The assessment data also yield scores for bottom-up empirically based syndrome scales that were derived by factor analysing problem scores for large samples of people who were assessed by the relevant kinds of informants. Like the DSM-oriented scales, the syndrome scales are scored quantitatively by summing their constituent items. Also like the DSM-oriented scales, the syndrome scales are displayed on profiles in relation to age-specific norms for each gender.

The ability to score DSM-oriented scales and empirically based syndromes from the same assessment instrument makes it easy to evaluate people in terms of both the top-down and bottom-up approaches. If an individual obtains scores in the borderline or clinical range on DSM-oriented scales, the DSM should be consulted to see whether criteria for DSM diagnoses are met.

Multi-informant assessment

Parallel forms are designed to obtain data from self-reports and reports by others who know the person being assessed. The data from the different respondents are scored on parallel scales. The scales are displayed on parallel profiles in relation to norms for ratings by each type of respondent. For example, scores obtained from adolescents' self-reports are displayed on profiles in relation to norms derived from self-ratings by large, representative samples of adolescents, separately for each gender. Parents' ratings of adolescents are displayed on profiles in relation to norms derived from ratings by large, representative samples of parents. And teachers' ratings of adolescents are displayed in relation to norms derived from ratings by large, representative samples of teachers.

Hand-scored profiles scored from each informant can be visually compared to identify consistencies and disparities with respect to the scores obtained from different respondents for specific items and scales. Computer software for scoring the forms prints sided-by-side displays of scores on problem items obtained from up to eight respondents. It also prints histograms that provide side-by-side comparisons of normed

DSM-oriented scale scores and syndrome scores from each respondent. This makes it easy to see at a glance whether any scale scores from any respondents are deviant and to determine whether the scales are consistently deviant across reports by different respondents. To provide a quantitative index of how well the different respondents agree, the software displays Q correlations between the ratings of problem items by each pair of respondents. To help users evaluate the levels of agreement, the 25th percentile, mean, and 75th percentile Q correlations are displayed for large reference samples of respondents.

Multi-cultural assessment

The instruments to be described in the keynote address have been translated into 69 languages. There are over 1,400 published reports of cross-cultural applications in 62 cultures. Comparisons of scores obtained by large, representative samples of children and youth in diverse cultures have shown that the mean problem scores from most of the cultures are remarkably similar, although some cultures have significantly lower or higher scores than most of the others. Work is now under way to perform multi-cultural factor analyses of data for over 60,000 children from 30 cultures to determine whether a single factor model fits all these cultures or whether multiple factor models are needed.

In summary, the keynote address will present a variety of important challenges facing assessment of psychopathology and adaptive functioning in the 21st century. It will also present practical ways to meet the challenges.

Viewing Organisations Through the Lens of Type:
Applications of the MBTI® Tool

Katherine W. Hirsh

Sandra Hirsh Consulting (MBTI®)

Katherine W Hirsh is a faculty developer at the University of Minnesota. As consultant and writer for Sandra Hirsh Consulting (MBTI®), Dr Hirsh was involved in team building and management training, as well as research, editing and writing. She is the co-author, with Elizabeth and Sandra Hirsh, of *Introduction to Type* and also *The MBTI® Teambuilding Program; A Leader's Resource Guide.*

Abstract

The MBTI® Tool is a self-report inventory designed to sort individuals according to their preferences on four dichotomous dimensions: extraversion-introversion, sensing-intuition, thinking-feeling, and judging-perceiving. Applying the lens of MBTI® type can help us to understand behaviour at work – both in terms of *what* we do as well as in terms of *how* we do those things we do. Through a set of four vignettes, one for each of the four dimensions of type, the application of the type lens is demonstrated. Each vignette illustrates the application of the type lens to a key work issue: problem-solving and conflict resolution through the lens of extraversion and introversion; leadership through the lens of sensing and intuition; stress through the lens of thinking and feeling; organisation change through the lens of judging and perceiving. In addition, guidance on when it is appropriate to apply the lens of type is provided.

Introduction

The MBTI® Tool is a self-report inventory designed to sort individuals according to their preferences on four dichotomous dimensions. The first of these dimensions, extraversion-introversion, reflects individual preferences in terms of energy. People with a preference for extraversion tend to be energised by the external world of people and things while people with a preference for introversion tend to be energised by the internal world of ideas. The second dimension, sensing-intuition reflects individual preferences for taking in information. People with a preference for sensing tend to take in information from the five senses in a precise and detailed fashion. People with a preference for intuition tend to take in information in an ad hoc manner with attention to patterns and associations. The third dimension, thinking-feeling, reflects individuals preferences for making decisions. People with a preference for thinking tend to make decisions that are rational and logical and they place a premium on fairness. People with a preference for feeling tend to make decisions that are individualised and personal and they place a premium on harmony. The fourth and final dimension, judging-perceiving, reflects individual preferences in terms of lifestyle organisation. People with a preference for judging tend to organise their lives around plans and commitments. People with a preference for perceiving tend to organise their lives around choices and spontaneity.

The lens of type helps us to understand behaviour at work both in terms of what we do as well as in terms of how we do the things we do. What follows is a set of vignettes designed to introduce you type as a lens for viewing organisations. The vignettes describe characteristics that tend to be associated with each of the eight MBTI® preferences in a work setting. Although an individual, team or organisation is more than the sum of each of the preferences, breaking behaviour down in terms of particular preferences provides a starker picture of the influences of the preferences at work and may better serve you in anticipating the strengths and challenges associated with particular type preferences. As you read these vignettes, keep in mind the fact that type is *not* the only source of variation in behaviour and that even individuals or teams with the same preferences will not have identical responses to a given situation.

Problem-solving and conflict resolution through the lens of extraversion and introversion

A good decision requires facts about the current situation, insight into future implications, a weighing of pros and cons, and a concern for values, especially as they impact people. Knowledge of MBTI® preferences gives us a window into what contributes to optimal decision making. To illustrate this, compare the problem solving styles of two imaginary teams, one with a preference for Extraversion and one with a preference for Introversion.

The extraverted team

This team approaches decision making at a rapid pace. They prefer to begin by talking about the problem with others, this talk is the trigger for internal information processing. Moreover, talking is typically the key means of communication – if possible this team would prefer to hear about a point of view rather than read about it. When the problem is being discussed this team will tend to focus on breadth, moving from topic to topic. In general this team is more focused on action than reflection.

The introverted team

This team approaches decision making at a measured pace. They prefer to begin by processing information internally, this internal processing is the trigger for talking with others. Talking is typically not the key means of communication – if possible this team would prefer to read about a point of view rather than hear about it. When the problem is being discussed this team will tend to focus on depth, digging deeply into each topic before moving on. In general this team is more focused on reflection than action.

Suppose each of these teams is facing budget reductions. They are tasked with deciding which of the many important programs they are running can be cut without damaging either team's ability to fulfil its mission. What will each team need to do in order to make a good decision? Or to put it another way, what might the decision look like without certain things being done?

To make a good decision both teams will need to:

- talk as well as reflect;
- think about the breadth of the problem as well as thinking about the problem in depth;
- move, but don't move hastily; and
- share their thoughts in discussion as well as documenting the decision-making process in writing.

Leadership through the lens of sensing and intuition

A good leader must share his or her values and beliefs, think deeply about issues, execute decisions in a timely manner and entertain alternative strategies and policies. Knowledge of MBTI® preferences gives us a window into what contributes to effective leadership. To illustrate this, compare the leadership styles of two imaginary leaders, one with a preference for sensing and one with a preference for intuition.

The sensing leader

This leader tends to draw on experience and precedence. For this leader, being pragmatic is key. He or she is more likely to use accepted leadership structures and strategies rather than expend effort on innovations. This leader is likely to be well grounded in the immediate needs of the team, what is called for right here and right now. He or she is likely to have a good command of the facts and details that underlie policies and procedures.

The intuitive leader

This leader tends to draw on insight and inspiration. For this leader being innovative is key. He or she is more likely to expend effort on creating new leadership structures and strategies rather than make use of accepted ones. This leader is likely to be focused on the future needs of the team, the potential and the possible. He or she is likely to have a good grasp of the patterns and relationships among policies and procedures

Suppose each of these leaders is tasked with creating a blueprint for a new holiday pay system. The system they construct must be in line with company policies, be fair and equitable, and meet the needs of team members with a variety of work situations and lifestyles. What will each leader need to do in order to persuade team members that the new system is beneficial? Or to put it another way, how might the team respond if information about the new system is not communicated?

To effectively sell their new systems these leaders will need to:

- share the relationship of the new system to the old – continuity as well as the potential the new system offers for innovation;
- give the nuts and bolts details as well as the model that under-girds the new system;
- allow space for 'the best of what is' and the 'best of what could be' to influence the system; and
- take the pulse of team at the present time as well as consider their future needs.

Stress through the lens of thinking and feeling

A good team member must respond to stressful situations with aplomb, working to prevent themselves and others being exhausted by the situation the team is facing. Knowledge of MBTI® preferences gives us a window into how a team member can contribute to reducing

team stress. To illustrate this, compare the responses to stress of two imaginary team members, one with a preference for thinking and one with a preference for feeling.

The thinking team member

The team member with a preference for thinking contributes to reducing team stress by offering an objective and dispassionate approach to difficult situations. This individual treats colleagues in an honest and direct manner. He or she typically can be counted on to respond to stressful situations systematically, taking in all points of view. Other team members turn to her or him for a realistic appraisal and a questioning attitude.

The feeling team member

The team member with a preference for feeling contributes to reducing team stress by offering an accepting and supportive approach to difficult situations. This individual treats colleagues in a thoughtful and concerned manner. He or she typically can be counted on to respond to stressful situations considerately, taking in all points of view. Other team members turn to her or him for an optimistic appraisal and a hopeful attitude.

Suppose both teams are faced with a tight, externally imposed deadline where all team members must put in extra hours. To complete the project team members need to work together and work to their strengths. What will each team member need to do in order to minimise the stress associated with meeting the deadline? Or to put it another way, how might the team look if no one acknowledged the potential for stress overwhelming the team?

In order not to overtax themselves team members need to:

- make an accurate yet upbeat evaluation of the situation;
- respond to setbacks with realism as well as encouragement;

- ask tough questions and listen supportively to answers; and
- treat stressed out colleagues in an even-handed yet personal fashion.

Organisation change through the lens of judging and perceiving

A flourishing organisation is responsive and reflective, leveraging the skills and abilities of its members, able to recreate itself in light of internal and external challenges, demands and prospects. Knowledge of MBTI® preferences gives us a window into how organisations undergoing change manage this effectively. To illustrate this, compare the approaches to change of two imaginary organisations, one with a preference for judging and one with a preference for perceiving.

The judging organisation

This organisation typically values stability. They tend to plan their response to change. Once a plan has been agreed they want to move forward by establishing goals and timetables for action and once goals and timetables are in place they prefer to stay the course. They are inclined to take a steady and scheduled approach to the implementation of new policies and procedures.

The perceiving organisation

This organisation typically values opportunity. They tend to examine possible strategies for change and construct plans on the fly. Even if they agree a plan at the outset, it is still likely that this will be adapted as the process unfolds, revising timetables and goals as necessary. They are inclined to take a flexible and spontaneous approach to the implementation of new policies and procedures.

Suppose each of these organisations is implementing a new system of recognition and reward. How can they plan for and acknowledge resistance to

the new system and work to ensure that the rewards have meaning for members of the organisation? Or to put it another way, what would the organisation look like if not attention was paid to individuals concerns regarding change?

To make the implementation of the new system go smoothly, these organisations will need to:

- recognise the need for the organisation to be predictable as well as a dynamic;
- plan for the future as well as dealing with contingencies as they arise;
- ensure that the implementation stays on track without ignoring difficulties that arise; and
- commit to a plan commit to staying with the process.

It is my hope that these vignettes have given you an idea of the richness available by viewing organisations through the lens of MBTI® type. I would like to conclude with a few suggestions for when it is appropriate to apply the lens of type.

- Do use the MBTI® to analyse job requirements.
- Don't base selection decisions on MBTI® type.
- Do use the MBTI® to analyse task demands.
- Don't equate preferences and skills and then base task allocation decisions on MBTI® type.
- Do use the MBTI® to investigate options and alternatives for task accomplishment.
- Don't allow MBTI® type to be used as an excuse for refusing an assignment or completing it poorly.
- Do use the MBTI® to discuss strategies for building on strengths and meeting developmental challenges.
- Don't fail to develop a team member based on his or her MBTI® type.

An Intelligent Way to Assess Emotional Intelligence

David R. Caruso

Harris McCully Associates Inc

David Caruso is a management psychologist and a Research Affiliate in Psychology at Yale University. He is the co-author of the Mayer, Salovey, Caruso Emotional Intelligence Test, and several empirical articles on emotional and general intelligence. David recently published *The Emotionally Intelligent Manager* with his colleague Dr Peter Salovey.

Abstract

The concept of EQ is neither new nor different from existing personality constructs or competency models. However, the concept of an emotional intelligence, or EI, is a new construct which is conceived of as a standard intelligence. It is this new intelligence which can be measured reliably, using objective methods of assessment and yields adequate discriminant and predictive validity.

Introduction

Passionate debate is not uncommon in our field. Consider, for example, two points of view regarding the concept of 'EQ', or emotional intelligence.

- 'EQ is twice as important as IQ. EQ redefines what it means to be smart.'

- 'EQ does not exist.'

I argue that both of these broad claims are incorrect, and suggest that the popular view of 'EQ' is misguided and that broad-based criticisms of the field are misdirected.

The concept of emotional intelligence is familiar to layperson and scientist alike due to the influence of a popular, 1995 trade publication that bore the title *Emotional Intelligence*. It is likely that the book's sales were due to a number of factors (see Mayer, Salovey, & Caruso 2000 for a review), but the reasons include the extraordinary claims the book made regarding emotional intelligence (or 'EQ' as many call it). It was claimed that EQ was 'twice as important as IQ' and that unlike IQ, EQ can be learned meaning that 'anyone can be smart'.

Unbeknownst to many, this populist approach was loosely based upon several modest scientific attempts to define an emotional intelligence (Salovey & Mayer 1990). These authors defined emotional intelligence as the ability to understand feelings in the self and others, and to use these feelings as informational guides for thinking and action (Salovey & Mayer 1990). However, the popular book expanded and stretched the definition of emotional intelligence well beyond the meaning and intent of the original concept.

Unfortunately for the field, there has often been little differentiation between these popular concepts and the scientific approach, even within the academic research community. Thus, researchers frequently confuse the original ability conception of emotional intelligence with so-called 'mixed models' of emotional intelligence which are non-cognitive in nature and mix various parts of personality into a single construct labeled EQ.

An emotional intelligence must meet the criteria for an intelligence, and it must also operate on and with emotion. An examination of intelligence and emotion should result in a more sophisticated understanding of the field, including a recognition of the following.

- EI is not EQ.
- EI can be conceptualised as an intelligence and is related to other intelligences.
- EI can be measured reliably and objectively.
- EI has discriminant validity (beyond other intelligences and traditional measures of personality).
- EI predicts important outcomes, at levels one expects for most psychological constructs.
- The popular notions of 'EQ' – defined as non-intellective factors – are better considered traditional competency or personality constructs.

Intelligence

Despite efforts spanning decades, psychologists have yet to agree upon a consensual definition of intelligence. At its core, intelligence refers to the ability to reason and learn. Perhaps Wechsler said it best when he wrote that intelligence is 'The…capacity to act purposefully, to think rationally, and to deal effectively with [the] environment.' (cited in Matarazzo 1972, p. 79).

The intelligence debate also rages over how many intelligence exist, whether it is a single, unified construct or consists of related, but independent abilities. One of the more useful frameworks within which to position this discussion is that of Carroll (1993) who proposed three levels, or strata, of intelligence. At the most global level is 'g' or general intelligence. At the second stratum are specific abilities such as fluid and crystallised intelligences. The first, or most basic stratum consists of specific, elementary cognitive tasks. Carroll's work is important to our field for a number of reasons, but chief among them is that he supported the notion that multiple intelligences exist and that they are worthy of study. In fact, he noted that 'It is the thesis of this book that there exist a substantial number of distinguishable and important mental abilities – as many as thirty or more.' (p. 27) He also advocated research to explore intelligence: '…initial goal of further research in cognitive abilities should be, in spite of pure science, to determine what cognitive abilities exist and can be measured…' (p. 689)

Emotion

Psychologists have been ambivalent about emotion. Consider one view from the 1930s which perceived emotion as antithetical to cognition and reason: '[Emotions cause] a complete loss of control. [There is] no trace of conscious purpose.' (Young 1936). An opposing viewpoint, emphasising the adaptive value of emotion is illustrated by Mowrer: 'The emotions are of quite extraordinary importance in the total economy of living organisms and do not deserve being put into opposition with 'intelligence'. The emotions are, it seems, themselves a higher order of intelligence.' (Mowrer 1960).

Many theories of emotion exist, and it is not necessary for the EI construct to adopt any single emotion theory. Instead, EI needs to consider the functions of emotion such as those proposed by Frijda (1986) or Plutchik (1980). Emotion:

- occurs due to a change in the environment;
- occurs automatically;
- comes on quickly;
- involves changes in physiology;
- can vary in intensity;
- creates changes in attention and thought;
- motivates certain behaviour;
- has a subjective experience (feeling);
- dissipates rapidly; and
- serves an adaptive function

The key is that emotions contain data and they send us a signal about what's going on in the world, or within ourselves. Emotions direct our attention and motivate us to engage in certain behaviours. These emotional signals heighten awareness in general, redirect our attention from whatever it is you were doing to an important event, and get you ready to act in some way. The most important aspect of emotions is to give us critical data about our interpersonal world. As Ekman (2003) notes, emotions serve many functions, but the most important function they serve is to prepare us, quickly, for critical interactions with other people.

Emotional intelligence: the Mayer-Salovey-Caruso model

The strategic use of emotions to help us reason and think, and the application of reasoning to emotions, is emotional intelligence. We'll now take a closer look at our approach to emotions and intelligence.

Peter Salovey and Jack Mayer first proposed the existence of an emotional intelligence in 1990 (Salovey & Mayer 1990). Over the intervening years, the approach has been modified as we've learned more about this ability (Mayer, Caruso & Salovey 1999). Our most recent model of emotional intelligence consists of four related abilities that work together as a process or an approach to decision-making, judgment, and leadership. The four abilities are:

- the ability to accurately perceive emotions;
- the ability to generate emotions and use them to influence thinking;
- the ability to understand the rules emotions follow; and
- the ability to manage with emotions to include emotions in our thinking and behaviour.

Let's examine each of these four emotional abilities.

Identifying emotions

This first EI ability is the ability to accurately attend to emotions – in yourself, in other people, and in the world around you. Paying attention to emotions is critical, but it's not enough. Our model begins with self-awareness and emotional awareness in general, but we stress the accuracy of judgment and perception.

Utilising emotions

The second ability is being able to create a certain feeling or emotion and

to use that emotion to help your thinking process.

Understanding emotions

Emotions have rules, just like the pieces of a chessboard. Another aspect of understanding emotions is the language of emotion. This ability also involves having a complex emotional vocabulary so that you can better understand and describe emotions.

Managing emotions

Since emotions contain data, to ignore them means that we are ignoring important information. This ability allows us to stay open to our emotions, no matter how uncomfortable they may be. But it also allows us to process the emotions and to leverage their information and their power to help us make better decisions and follow a wiser course of action.

Emotional intelligence: other models

The ability approach to emotional intelligence is similar to other proposed intelligences such as intrapersonal and interpersonal intelligences (Gardner 1983), social intelligence (Riggio, 1986; Thorndike 1920) and practical intelligence (Sternberg & Caruso 1985). Emotional intelligence overlaps each of these concepts, but is narrower and more focused on emotions per se.

There are two alternative conceptions of EI which my colleagues and I have termed mixed models. For example, a dissertation on psychological well-being consisting of a broad range of personality traits became the basis for a test of 'emotional intelligence' (Bar-On 1997). This model of EI, or EQ consists of five trait categories: intrapersonal (self-regard, emotional self-awareness, assertiveness, independence, self-actualisation), interpersonal (empathy,

social responsibility, interpersonal relationship), adaptability (reality testing, flexibility, problem solving), stress management (stress tolerance, impulse control, and general mood (optimism, happiness).

The other populist notion of emotional intelligence defined it as a set of leadership competencies, drawn from decades of competency modeling research and application (Goleman 1998). Goleman defined an emotional competence as an acquired or learned skill, and one that is based upon emotional intelligence. This competency approach to EI included five clusters of skills: self-awareness (emotional awareness, accurate self-assessment, self-confidence); self-regulation (self control, trustworthiness, conscientiousness, adaptability, innovation); motivation (achievement, commitment, initiative, optimism); empathy (understanding others, developing others, service orientation, diversity, political awareness); and, social skills (influence, communication, conflict management, leadership, change catalyst, building bonds, collaboration/cooperation, team capabilities). In later work, the five clusters were reduced to four (self-awareness, social awareness, self-management, and social skills).

Measuring emotional intelligence

Trait measures of EI

Self-report tests ask a person to indicate how they feel, or how they evaluate themselves. An example of a self-report test for EI is the Bar-On EQ-i. This test asks people to report on their self-perceptions regarding aspects of their personality such as how happy they are, how assertively they act, and whether they believe that they are emotional aware. The problem with using this method to measure EI is that

a person's self-report of their skills, abilities, or intelligence is not a good predictor of their actual ability. Studies comparing the EQ-i with an ability measure of EI typically show correlations ranging between .00 and .30 (Brackett & Mayer 2003). At the same time, this self-report test correlates highly with traditional measures of personality.

Competency measures of EI

360-degree assessments typically obtain ratings from a number of sources: the target person, the target's peers, customers, direct reports, and supervisor. The Emotional Competence Inventory (ECI) is such a 360-degree assessment. Items on the ECI were drawn from three sources: the 1991 Self-Analysis Questionnaire, a rating survey designed to measure management competencies of MBA students; a listing of general competencies; and, additional competencies listed in Goleman's 1998 book. A 360-degree assessment suffers from the same problem as self-report: both are poor predictors of a person's actual ability level (Borkenau & Liebler 1993).

Other measures which may assess EI

Emotion recognition has been measured in the past, by tests such as the PONS (Rosenthal et al. 1979). Emotion regulation strategies and their effectiveness have been widely studied (see, for instance, Gross & John 2002; Thayer 1996).

Some intelligence researchers claim that EI is rooted in Binet and Wechsler's work, noting that Wechsler assumed that the variance not captured by intelligence scales was due to non-intellective factors. Further, Kaufman and Kaufman (2001) recently argued that some of Wechsler's tasks measure

aspects of EI, such as comprehension and picture arrangement. However, such analyses miss the point in that the ability approach to emotional intelligence is not 'non-intellective'.

Ability measures of EI

An ability test measures a person's capacity to solve certain kinds of problems, and their problem-solving performance is compared to some type of reference group. Early ability measures often suffered from poor reliability (consistency of measurement) and ambiguous factor structure (the abilities it measures). The newest such ability measure, the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), appears to have resolved many of these issues. The MSCEIT measures a person's emotional intelligence abilities by asking the person to solve a series of emotional problems. For example, the ability to identify emotions accurately is measured by viewing a face and indicating what emotions are expressed by that person.

The MSCEIT

The MSCEIT includes 141 items. These items are grouped into eight different problem sets, or tasks. There are two tasks that measure each of the four emotional intelligence abilities. The MSCEIT can be scored using one of two scoring methods: general consensus scoring is based upon the idea that emotions contain information about important interpersonal events, and that the responses of a large group of people can best define better, and worse, answers. The second method, expert consensus, employs the answers of emotions researchers to determine the best answers to MSCEIT questions. These two scoring methods converge, suggesting that there are indeed better and worse answers to MSCEIT items.

The MSCEIT offers total and four specific EI ability scores with adequate reliability. Confirmatory factor analyses support the four-ability model of EI. Further, the MSCEIT is relatively independent of traditional personality measures and is moderately related to analytical ability (Mayer et al. 2001, 2003).

Unlike the bold claims for EQ, the MSCEIT does not predict 80 percent of the variance for any outcome! Its levels of prediction are what one would expect for most psychological measures. It is not a necessary ingredient for effectiveness, nor is it the most important factor in 'success'. The MSCEIT does predict outcomes such as mental health, pro-social behaviour, aggressiveness, and effective team work.

Being smart about emotional intelligence

Emotional intelligence is a young field and has not yet achieved the level of sophistication of general intelligence and IQ testing. Indeed, it may never do so. But emotional intelligence is a promising, new concept and worthy of study. Researchers and practitioners alike should be smart about their approach to EI: to be reasonably skeptical of the field, and to clearly differentiate between the ability concept of EI and various naïve concepts of EQ.

References and further reading

Bar-On, R. (1997). *Bar-On Emotional Quotient Inventory: A measure of emotional intelligence*. Toronto, ON: Multi-Health Systems, Inc.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality & Social Psychology 65*, 546–553.

Brackett, M., & Mayer, J.D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin 29*, 1147–1158.

Carroll, J.B. (1993). *Human Cognitive Abilities: A survey of factor-analytical studies*. New York: Cambridge University Press.

Caruso, D.R. & Salovey, P. (2004). *The Emotionally Intelligent Manager*. San Francisco: Jossey-Bass.

Ekman, P. (2003). *Emotions Revealed*. New York: Times Books.

Frijda, N.H. (1986). *The Emotions*. New York: Cambridge University Press.

Gardner, H. (1983/1993). *Frames of Mind: The theory of multiple intelligences*. (10th Anniversary ed.). New York: Basic Books.

Goleman, D. (1998). *Working with Emotional Intelligence*. New York: Bantam.

Gross, J.J., & John, O.P. (2002). Wise emotion regulation. In L.F. Barrett & P. Salovey (eds), *The Wisdom in Feeling: Psychological processes in emotional intelligence* (pp. 297–318). New York: Guilford Press.

Kaufman, A.S., & Kaufman, J. C. (2001). Emotional intelligence as an aspect of general intelligence: What would David Wechsler say? Emotion, 1, 258–264.

Matarazzo, J.D. (1972). *Wechsler's Measurement and Appraisal of Adult Intelligence* (5th ed.). New York: Oxford University Press.

Mayer, J.D., Caruso, D.R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence 27*, 267–298.

Mayer, J.D., DiPaolo, M.T., & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: a component of emotional intelligence. *Journal of Personality Assessment 54*, 772–781.

Mayer, J.D., Salovey, P., & Caruso, D. (2000). Models of emotional intelligence. In R.J. Sternberg (ed.), *The Handbook of Intelligence* (2nd ed., pp. 396–420). New York: Cambridge University Press.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitaraneos, G. (2001). Emotional intelligence as a standard intelligence. *Emotion 1*, 232–242.

Mayer, J.D., Salovey, P., Caruso, D.R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion 3*, 97–105.

Mowrer, O.H. (1960). *Learning Theory and Behaviour*. New York: Wiley.

Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. New York: Harper and Row.

Riggio, R.E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology 51*, 649–660.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). The PONS Test. Baltimore, MD: Johns Hopkins University Press.

Salovey, P., & Mayer, J.D. (1990). Emotional intelligence. Imagination, Cognition, and *Personality 9*, 185–211.

Sternberg, R.J. & Caruso, D.R. (1985). Practical modes of knowing. In E. Eisner (ed.) *Eighty-Fourth Yearbook of the National Society for the Study of Education (Part II)* (pp. 133–158) Chicago: University of Chicago Press.

Thayer, R.E. (1996). *The Origin of Everyday Moods: Managing energy, tension, and stress*. New York: Oxford University Press.

Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine 140*, 217–235.

Young, P.T. (1936). *Motivation of Behavior*. New York: John Wiley & Sons. Quotes from pp. 457–458.

Quality of Performance and Change-Sensitive Assessment of Cognitive Ability
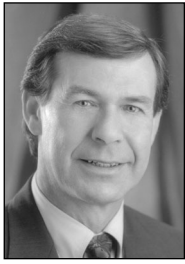
Gale H. Roid

Vanderbilt University

Gale Roid is the Dunn Professor of Educational and Psychological Assessment, Department of Special Education, Peabody College of Vanderbilt University, Nashville Tennessee. Author of many well-respected publications including *Stanford-Binet Intelligence Scales Fifth Edition*, Professor Roid has also co-authored the *Stoelting Brief Nonverbal Intelligence Test (S-BIT)* and the *Leiter International Performance Scale –Revised.*

Abstract

Growth and change are fundamental processes in human development. Recent trends in education, psychology, medicine, and other fields have stimulated renewed interest in methods of measuring developmental growth and change (Collins & Sayer 2001). Educators and governmental agencies have recently placed greater emphasis on the importance of growth in reading as foundational to student learning (Torgesen 2002). A major review of the regulations is underway in the United States, concerning funding of special education in public schools (e.g., President's Commission on Excellence in Special Education 2002) and the emphasis has been on 'continuous progress' assessment and accountability for student learning (e.g., multiple testing during the school year). Extensive research is underway on the developmental time tables in antisocial behaviour (Bauer & Estell 2001). Highly sophisticated multivariate statistical models are being developed for assessing change in medical programs directed at drug-use prevention, depression recovery, and head-injury recovery (Collins & Sayer 2001). Measures of growth are also needed in monitoring the progress of infants born prematurely (Krishnakumar & Black 2001; Roid & Sampers 2004). When cognitive performance decreases rather than increases, as in the elderly (e.g., memory function), measures sensitive to change in the negative direction are needed (Roid 2003b).

Definitions

'Growth' refers to any incremental improvement in cognitive functioning, however small. Growth is most obvious with repeated, individual (longitudinal) testing. Increments of growth are analogous to the changes in performance noted across age groups, from birth to adulthood, as measured by growth curves of test scores. 'Change', in the context of the current paper, means any increment of improvement, decline or recovery in cognitive functioning. This change may be due to a variety of causes, including typical cognitive development, injury or illness, or response to treatment or intervention. 'Change-sensitive assessment' refers to any evaluation, based on test scores and other information about an individual that is collected or studied at two (or more) points in time and used to evaluate growth or change. Change-sensitive assessments are particularly helpful in evaluating learning capacity, response to intervention, effectiveness and appropriateness of treatment and general tracking of growth or change in an individual across time. 'Quality of performance' methods refer to testing or observational procedures that are designed to identify small increments of difference in the quality (not just quantity or presence/absence) of actions, behaviour, performances, or products created by the target individual being assessed. Quality-of-performance measures help in the identification of borderline, or mild, developmental delays because children may obtain 'milestones' (behaviour occurring at the expected age such as walking by age 1 year) but with unexpected quality or atypical characteristics.

The role of item response theory in change-sensitive assessment

Major measurement tools for building change-sensitive assessments have been

developed through research on item response theory (IRT) models. IRT models (Lord 1980) are a large family of mathematical models used to analyse test items, develop collections of items, create scales, and produce test scores for examinees. Following decades of research on IRT models, Embretson (1996) recently asserted that the models had reached such an acceptable level of scientific verification that they should replace classical test theory (e.g., Gulliksen 1950) as the 'new rules of measurement' in psychology and education.

The version of item response theory that I have applied most often to individually-administered tests in psychology and education is the Rasch model, named for the Danish mathematician, Georg Rasch (Rasch 1966, 1980). Rasch proposed that performance on a test can be predicted from the ability (A) of the examinee and the difficulty (D) of the item. Embretson (1996) praised the advantages of the additive model, based on a fundamental tenet of measurement theory – additive decomposition – in which two parameters are related to a third variable (e.g., a measurement scale) by an additive (subtractive) relationship. Embretson said, 'In the Rasch model, additive decomposition is achieved; the log odds that a person endorses or solves an item is the simple difference between his or her trait level…and the item's difficulty…' (p. 348).

The SB5 and Rasch analysis

The Rasch model was used in several ways and in several stages of the development of Stanford-Binet Intelligence Scale, Fifth Edition (SB5, Roid 2003a). Some of the important uses of the model and its advantages included item analysis, item calibration, and development of change-sensitive scores (CSS) for each of the major summative scores (4 IQ scores and 5 cognitive-factor indexes). With the Rasch model, both item difficulty and examinee ability are scaled in the same measurement metric. Difficulty calibrations and ability are initially estimated by computer programs and the values appear as normal-curve z-scores (called 'logits' or log units, Lineacre & Wright 2000), ranging from minus 4.0 to plus 4.0. For better interpretability, the difficulty values for each SB5 item and the resulting CSS (estimates of examinee ability) were converted to the W-scale developed by Woodcock and Dahl (1971). The W-scale transforms the initial logit values by centering them at 500 and using a special expansion factor of 9.1024, developed by Woodcock and Dahl. Thus, the CSS scale and item difficulty scale for SB5 ranges from approximately 425 for 2-year-old children to 525 for adults, with a central value of 500 located at the mean performance level of children 10 years, 0 months of age (beginning fifth grade approximately). The CSS scale and item difficulty have a criterion-referenced interpretation based on age equivalence, task characteristics (e.g., complexity of the SB5 items), and overall sequence of cognitive development suggested by the scale. As a child progresses upward on the scale, he or she is capable of mastering increasingly complex tasks and solving increasingly challenging problems. This progress mirrors the development of the brain, the growth of academic competencies, and the accumulation of general knowledge. In addition to norm-referencing, where the child is compared to peers of the same age, the CSS scale allows for criterion-referencing to task complexity, and age-related milestones such as the achievement of reading fluency or the

various stages in mathematical competence.

CSS scores are available for Full Scale, Nonverbal, Verbal, and Abbreviated IQ and for the five cognitive factors from the Cattell-Horn-Carroll theory (Carroll 1993; Horn & Cattell 1966; Flanagan 2000). When these CSS scores are plotted across age groups, using cross-sectional (not longitudinal data), the classic 'growth curve' shapes are evident. The cognitive-factor curves increase from the early childhood years through the early twenties, and, then, depending on the cognitive factor being measured, begin to show declining scores in older age groups. Memory CSS scores show the most rapid decline across elderly age groups, perhaps due to the emergence of dementia, Alzheimer's disease, etc. An exception is the crystallised (General Knowledge and Vocabulary) ability factor which shows continuing improvement into the late 50s among older adults.

Rasch growth scores in other tests

Previous applications of the Rasch model were made in the Woodcock-Johnson Psychoeducational Battery, Revised (Woodcock & Johnson 1989), the Toddler and Infant Motor Evaluation test (TIME, Miller & Roid 1994), in the Leiter International Performance Scale, Revised (Leiter-R, Roid & Miller 1997), and in the new Merrill-Palmer Developmental Scales, Revised (MP-R, Roid & Sampers 2004). These instruments and the 'growth scores' in them have generally been received positively by professionals working with disabilities or developmental delay. The potential is great for detailed tracking of growth or change across time, and the interpretive power of criterion-referenced scales such as CSS. A striking consistency across national

standardisations and across test developers has begun to emerge when the CSS or Growth or W-scale scores are compared across cognitive batteries such as the SB5, the WJ-R, the TIME, the Leiter-R, and the MP-R. The Rasch-based scores on each of these tests have been anchored to the value of mean score of children, age 10 years, 0 months (or, in the case of the MP-R, at 460 for age 4 years, 0 months). Theoretically, the ends of each scale could them depart in various ways across batteries. However, excellent consistency has been achieved across these diverse test batteries (e.g., consistency of 425 as a value at age 2).

Quality of performance and change-sensitive measurement

Another important advance in measurement that makes change-sensitive assessment possible is the development of instruments sensitive to the quality of the individual's performance. Rather than simple counts of the number of correct responses or the number of behavioural milestones achieved on schedule (e.g., early vocabulary before age 1, walking at about age 1, learning to read by age 8 or 9), the unique quality of responses can be observed and recorded. Examples of performance quality assessments are listed below and will be described in more depth in the presentation:

Movement quality in infants and toddlers

For example, quality of movement in infants and toddlers was studied as part of the development of a test called the Toddler and Infant Motor Evaluation (TIME, Miller & Roid 1994). Detailed observations of children with both typical and atypical motor development

were taken and detailed illustrations of children in various movement positions were drawn. Examiners using TIME can observe a child moving from a prone position to standing in a 12-month old child, for example. Observations are made every 5 seconds and recorded on the test. The pattern of the movements, not simply the final position (standing) is important in identifying mild and moderate developmental delays. The child should roll over, use hands, arms, and knees to lift himself or herself from the floor, and then use one leg (with perhaps a hand on a chair) to move to a standing position, in the typical pattern. Odd positions of hands, arms, back, legs, etc., may indicate atypical movement. Thus, the quality of the movement is assessed with the TIME system. Miller and Roid (2003) used a sequence comparison method (Jackson 1990; Sellers 1974) to compare typical patterns (stored in a computer program) to the patterns observed in typical and atypical children, with excellent discrimination. Details of the method and research will be discussed in the paper.

Quality of cognitive performance on the SB5

Guidelines for interpreting the Stanford-Binet, Fifth Edition (Roid 2003a) include recommendations for the qualitative assessment of child performance on certain subtests and items. For example, the quality of fine motor movement exhibited by children while assembling the pieces of the Form Board or Form Patterns tasks can vary from exceptional, typical, to unusual and atypical movement, modes of grasping the pieces, etc. Most striking, the strategies used by the child to sort the picture chips in the Verbal Fluid Reasoning task are very interesting. The task is to sort the chips into groups of three. Some children only use very concrete categories such as color.

Others use functional categories such as 'writing utensils', or 'play equipment', revealing the quality of their developmental level of thinking. Such qualitative details can be lost if the tasks are not designed to allow their observation or if examiners do not attend to them.

Play-based quality of performance measures: the MP-R

The new revision of the classic Merrill-Palmer Developmental Scales (Stutsman 1948; Roid & Sampers 2004) includes several toy-based tasks that tap the quality of infant and child cognitive and fine-motor abilities. A 'spin toy' reveals the infants quality of hand movement and hand–eye coordination. The 'problem box' (a clear plastic box with interior shelves into which a small toy is inserted with the task to extract the toy) reveals many problem-solving (fluid reasoning) strategies in children. Some children shake the box and pound it on the floor or table. Others try to reach into the small openings in the box. Others discover the bottom 'flap' and open it to extract the toy. These toy-based tasks provide great richness of quality performance assessment, and provide indicators of advanced, typical, or delayed/atypical performance for purposes of early identification of developmental disabilities.

Assessment of essay writing in school children

Data on 10,000 students in the public schools of the State of Oregon (USA) were studied by Roid (1994). Essays from these students were graded using a 6-point, analytical trait method of performance assessment with substantial inter-rater reliability. The ratings produce 6 trait scores for each essay (each student) on dimensions such as quality of word choice,

grammar and mechanics, creative expression ('voice'), organisation, etc. Roid (1994) used cluster analysis to identify groups of students with similar patterns of trait scores and found groups that had high creativity versus poor mechanics of writing.

Assessment of fluid reasoning in infants

One challenging area of assessment is identifying the quality of fluid reasoning in children under the age of 2 years. Prior to work on the Merrill-Palmer revision (MP-R), few published tests provided standardized measures of infant reasoning, except the Bayley Scales of Infant Development and a few others. Also, existing measures did not have 'change-sensitive scores' or quality-of-performance items as in the MP-R. Now, the MP-R provides a downward extension of Woodcock's W-scale down to a value of approximately 327 for age 1 month, based on cognitive play-based tasks, observations of eye-movements in tracking toys, etc. These findings will be discussed in the context of the challenge of early assessment of fluid reasoning.

Change sensitive assessment and the evaluation of cognitive delays in premature infants

Assessing premature infants is an area of important advancement promised by the development of change-sensitive scores and methods of measuring quality of performance. As part of a federally-funded research program, the developers of the Merrill-Palmer Developmental Scale, Revised (MP-R, Roid & Sampers 2004) have begun to study the problem of using 'age corrections' on developmental scales. Because premature infants are often

born 4 to 8 weeks prior to typical gestation, scores on their future developmental tests are often 'corrected' by using norm tables one or two months lower than the chronological age (measured from birth) for those infants. Lems, Hopkins, & Samsom (1993) suggested that a full correction for children in the first 6 months of life may overestimate the child's score and that a lack of correction will underestimate the child's abilities. The correction may mask a true delay. When, exactly, does the correction diminish and by what magnitude? Aylward (2002; 1997) suggests that the degree of correction to accurately predict outcomes of premature infants will require an algorithm based on the age of the infant, background risk factors, and, importantly, the domain of cognitive, motor, or language behaviour being assessed. Recent research using the new MP-R will be reviewed to show progress made in examining the age correction dilemma.

Summary

More than a decade of research has been conducted to study and develop instruments sensitive to developmental growth and decline in cognitive functioning. Many applications to important assessment problems in education, psychology, medicine, special education, and infant evaluation have been discussed. Many challenges remain for future researchers, including continuing studies using true longitudinal research designs, experimental studies of premature infants 'catching up', and studies of early-emerging cognitive abilities such as fluid reasoning. Possible technology developments in the future may be promising, such as use of personal ('palm') data-collection devices to test children more frequently across time. Finally, one of the promising advantages of change-sensitive assessment is the ability to show parents of children with special needs that their children are making progress predicted by the patterns of documented growth curves.

References

Aylward, G.P. (2002). Methodological issues in outcome studies of at-risk infants. *Journal of Pediatric Psychology 27(1)*, 37–45.

Aylward, G. P. (1997). Conceptual issues in developmental screening and assessment. *Journal of Developmental and Behavioral Pediatrics 18(5)*, 340–349.

Bauer, D. J., & Estell, D. B. (2001). Cluster analysis of developmental profiles: Relations between trajectories of aggression and popularity over adolescence. In L. Collins & A. Sayer (eds), *New Methods for the Analysis of Change* (pp. 385–387). Washington, DC: American Psychological Association.

Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. New York: Cambridge University Press.

Collins, L. M., & Sayer, A. G. (eds) (2001). *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Embretson, S.E. (1996). The new rules of measurement. *Psychological Assessment 8*, 341–349.

Flanagan, D. (2000). Wechsler-based CHC cross-battery assessment and reading achievement. *School Psychology Quarterly 15*, 295–329.

Horn, J.L., & Cattell, R.B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology 57*, 253–270.

Gulliksen, H. (1950). *Theory of Mental Tests*. New York: Wiley.

Jackson, D. F. (1990). The use of sequence analysis algorithms in research on computer-assisted problem solving strategies. Paper presented at the meetings of the American Educational Research Association, Boston, April.

Krishnakumar, A., & Black, M. (2001). Estimating cognitive growth curves from environmental risk factors: Mediating the role of parenting and child factors. In L. Collins & A. Sayer (eds), *New Methods for the Analysis of Change* (pp. 414–415). Washington, DC: American Psychological Association.

Lems, W., Hopkins, B., & Samson, J. F. (1993). Mental and motor development in preterm infants: the issue of corrected age. *Early Human Development 34*, 113–123.

Lincacre, J.M., & Wright, B.D. (2000). WINSTEPS v. 3.00: *Rasch Item Analysis Computer Program Manual*. Chicago: MESA Press.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.

Miller, L.J. & Roid, G.H. (1994) *The T.I.M.E. Toddler and Infant Motor Evaluation*. San Antonio, TX: The Psychological Corporation.

President's Commission on Excellence in Special Education. (2002). *A New Era: Revitalizing Special Education for Children and their Families*. Washington, DC: U. S. Department of Education.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology 19*, 49–57.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.

Roid, G. H. (2003a). *Stanford-Binet Intelligence Scales, Fifth Edition Examiner's Manual*. Itasca, IL: Riverside.

Roid, G.H. (2003b). *Stanford-Binet Intelligence Scales, Fifth Edition Technical Manual*. Itasca, IL: Riverside.

Roid, G.H. (1994). Patterns of writing skills derived from cluster analysis of direct-writing assessments. *Applied Measurement in Education 7*, 159–170.

Roid, G.H., & Miller, L.J. (1997). *Leiter International Performance Scale – Revised Manual*. Wood Dale, IL: Stoelting.

Roid, G.H., & Sampers, J. (2004). *Merrill-Palmer Developmental Scale – Revised Manual*. Wood Dale, IL: Stoelting.

Roid, G. H., & Woodcock, R. W. (2000). Uses of Rasch scaling in the measurement of cognitive development and growth. *Journal of Outcome Measurement 4(2)*, 579–594.

Sellers, P. H. (1974). On the theory and computation of evolutionary distances. *SIAM Journal of Applied Mathematics 26*, 787–793.

Stutsman, R. (1948). *Merrill-Palmer Scale of Mental Tests*. Wood Dale, IL: Stoelting.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology 40*, 7–26.

Woodcock, R. W., & Dahl, M. N. (1971). *A Common Scale for the Measurement of Person Ability and Test Item Difficulty*. (AGS Paper No. 10). Circle Pines, MN: American Guidance Service.

Woodcock, R.W., & Johnson, M.B. (1989). *Woodcock-Johnson Tests of Cognitive Ability – Revised*. Chicago: Riverside.

Wright, B.D. & Stone, M.H. (1979). *Best Test Design*. Chicago: Mesa Press.

Concurrent Papers

Applying the Mayer-Salovey-Caruso Approach to Emotional Intelligence

David R. Caruso

Harris McCully Associates Inc

David Caruso is a management psychologist and a Research Affiliate in Psychology at Yale University. He is the co-author of the Mayer, Salovey, Caruso Emotional Intelligence Test, and several empirical articles on emotional and general intelligence. David recently published *The Emotionally Intelligent Manager* with his colleague Dr Peter Salovey.

Abstract

How are you? It's an important question. It's important because you feel a certain way right now, and these feelings are directing your thoughts in certain areas. You feel this way for a reason, and understanding the reasons and what's happening to you and around you can help you determine how the feelings will change. Whether you are feeling great or lousy is important because these feelings are sending you a message. You can ignore the message at your own peril, or attend to the emotional communication to make better decisions.

Emotions are complex and important, and above all else, they can be *intelligent*. With emotions playing a critical role in our decisions, and actions, wouldn't it be helpful to be able to objectively measure our emotions in some way? This is what the Mayer, Salovey, Caruso Emotional Intelligence Test (the MSCEIT, which we pronounce 'Mess-keet') does. The MSCEIT is like an IQ test for emotions. It is a powerful and sophisticated tool, which measures a person's actual emotional skills.

This ACER workshop covers the following topics:

- About Emotional Intelligence
- Measuring Skills
- The MSCEIT
- MSCEIT Applications

About emotional intelligence

Emotions are a complex and sophisticated form of information. This form of reasoning has been known as *emotional intelligence*. While there are many approaches to *emotional* intelligence, the approach I'm talking about here is the original, scientific conception. It's based upon the decade of research and theorising by psychologists Jack Mayer and Peter Salovey, the originators of the theory of emotional intelligence. This is an *ability* model of emotional intelligence, which defines emotional intelligence as the *ability* to reason with and about emotions. Emotional intelligence combines feelings with thinking, and thinking with feeling, and can be described as four related, but different, abilities.

- **The ability to identify emotions.** Emotions contain information, or data. The ability to identify emotions allows you accurately to recognise how you and those around you are feeling.

- **The ability to use emotions.** This is the ability to generate emotions, and to use them in cognitive tasks such as problem-solving and in creativity.

- **The ability to understand emotions.** This includes the ability to understand complex emotions and emotional 'chains' – how emotions transition from one stage to another.

- **The ability to manage emotions.** This is the ability that allows you intelligently to integrate the data of emotions in yourself and in others in order to devise effective strategies that help you achieve positive outcomes.

The MSCEIT is based on this intelligent understanding of emotion.

Ways to measure skills

There are many ways to measure people's skills. Consider an example from American baseball – not the professional sport, but from a recent little league experience I had with

children aged 10 to 12 years. Coaches evaluate players during spring try-outs. Each batter gets 5 pitches. A miss gets 0 points, making contact yields 1 point, and a solid hit is 2 points, resulting in a possible score of 0 to 10. Danny, a 12-year-old, steps into the batting cage with a bit of bravado and swagger.

He gave the coach behind the pitching machine a bored, disinterested look, and took a practice swing. He proceeds to foul off a pitch (1 point each), hits a blooper on 1 pitch (1 point), a weak foul ball on another pitch (1 point), and swings and misses on the last 2 pitches (0 points). It seems clear that Danny's score is about a 4. However, the two coaches who are filling out the evaluation form put their heads together and proceed to give Danny a rating of 7.5. To a quizzical look, one of them replies 'You know Dan, the big guy? This is his kid.' Dan is another coach, one of the league commissioners, and there was no way that his coaching colleagues were going to give a son of Dan's a '4' rating.

And how did Danny evaluate his own performance? When Danny was asked about his performance by one of this friends, he said, 'Well, I was pretty good. I'm a great hitter anyway. I don't know, I'd say a 7 or 8, around there.' As the season played out, Danny's hitting was, indeed, not a whole lot better than a 4.

Measuring EI

A MSCEIT score is not based upon self-perceptions, nor is it based upon your reputation or other, political factors. It's based on how many hits and misses you get when you step up to the plate and determine how people feel, understand the cause of emotions, and determine optimal emotional strategies.

Many of us are familiar with people who are extremely self-aware, and aware of other people. The problem is that their awareness is faulty. This is a person who buttonholes you as you are dashing out of the door to an important meeting, oblivious to your look of near panic. Of course, it might be a critical point your colleague wants to raise with you, except that he launches into a broad discussion regarding his upcoming weekend plans – that don't involve you.

That's why it is extremely important that we are able to objectively measure emotional skills – through the use of an ability test. You have to find out what the person's *actual* skills are. Is it also important to understand how a person views his or her skill set? Absolutely! Is it of value to determine how others view a person? Of course. The MSCEIT is but one way of measuring something.

The MSCEIT

The MSCEIT measures the four core emotional abilities defined in the Mayer-Salovey model.

If you want to measure the ability to *accurately* identify how people feel, one way to do so is by asking the test taker what emotions are being expressed in a photograph of someone's face. For example, if you show a photo of a person displaying mild sadness, and the test taker selects an answer indicating that the person is feeling a bit happy and somewhat surprised, then such an answer is considered incorrect.

The MSCEIT includes many such tasks, each designed to test a person's emotional ability. The chart below lists the 8 tasks on the MSCEIT.

Ability	Test sections	Question types
Identifying	Faces	Identify subtle emotions in faces.
	Pictures	Identify emotions in complex landscapes and designs.
Using	Facilitation	Knowledge of how moods impact thinking.
	Sensations	Relate various feeling sensations to emotions.
Understanding	Changes	Multiple choice questions about how emotions change over time.
	Blends	Multiple choice emotion vocabulary definitions.
Managing	Emotion Management	Indicate effectiveness of various solutions to internal problems.
	Emotional Relations	Indicate effectiveness of various solutions to problems involving other people.

Scoring the MSCEIT

The point has been made that the MSCEIT is different from other tests. The types of questions are different, the abilities it measures are different, and the way that it's scored is also different.

There are two answer keys for the MSCEIT. The first key (general consensus) is based upon the consensus of thousands of people to converge upon better and worse answers. General consensus scoring works because emotions communicate information about people.

The second key (expert scoring) is based upon the answers of a panel of emotions experts. This is somewhat similar to the process used by standard IQ tests.

You also have the option of generating a report using norms based on age, ethnicity, and/or gender. EI does vary somewhat across these groups, and your scoring decision will vary depending upon your application. Not sure which way to go? My preference is to use Expert scoring with no corrections. It's clean and straightforward.

Training EI

Unlike IQ, EQ can be trained, right? Not so fast: it depends upon how you define and measure EI. If you define EI as a collection of traits such as assertiveness or leadership, the answer is 'perhaps'. If you define EI as an intelligence, as my colleagues and I do, the answer is either 'no' or 'we're not yet sure'.

We have yet to examine training efficacy for EI programs. However, if we refer to literature on intervention and education programs in general, one can conclude that intelligence is not going to be taught, but that remedial strategies, specific content, and compensatory strategies can be.

The practical take-away is for the design of training and program effectiveness studies of EI. The MSCEIT, as an ability test, is not an appropriate pre-post test measure. The outcomes which we would expect from an ability-based approach to EI should be based on certain specific skills, knowledge or behaviours. The behaviours may include conflict resolution, quality of communications, pro-social behaviours, and management of others.

MSCEIT applications

We use the MSCEIT in a number of ways. Here are some examples of some common MSCEIT applications:

* selection and promotion;
* career development;
* executive coaching and leadership development;
* counselling and therapy; and
* seminars and workshops.

Let's look at a few of these in greater detail.

Using the MSCEIT in executive coaching

Advantages of using the MSCEIT for executive coaching

The MSCEIT provides a unique look at a person's management and leadership skills. While most senior executives find management assessment to be of interest and value, they are often not surprised by their assessment results. Certainly, the results are useful, but the MSCEIT consistently provides information of a *different* sort. In fact, when I get to their MSCEIT results, it is common for the client to say something like 'That was the test that was a little different. What was that all about?'

How to use the MSCEIT for executive coaching

The MSCEIT, like all tests, can help you to develop questions, or hypotheses,

about a client. Each of the four, key MSCEIT scores can generate discussion around key coaching objectives.

MSCEIT Score	Questions to ask
Identifying	Does the person 'read' others well?
Using	Do they connect emotionally? Are they idea oriented?
Understanding	Do they perform adequate what-if analyses regarding people?
Managing	Are they effective decision-makers?

In addition, executives readily understand, and can apply, their MSCEIT results in a broad fashion through the use of the Emotional Blueprint.

Example of using the MSCEIT for executive coaching

Jerry was an operations manager for a major Wall Street firm. He was asked to re-locate most of his staff from NYC to a new building across the river in New Jersey, about a 10 minute ferry ride away. Most of his staff lived in New Jersey and welcomed the move. Jerry, and a few of his staff, were remaining in NYC.

The move itself went well, but there arose a number of unusual personnel problems in the following weeks. The problems consisted mainly of complaints at first. Jerry patiently listened to these complaints, addressed them, and understood their cause. As each problem was addressed, and resolved, a new one appeared. The problems increased in frequency, and began to have a noticeable impact on the group's productivity. At this point, Jerry was referred for executive coaching to help him resolve these issues.

Jerry's MSCEIT scores were as follows.

MSCEIT score	Low				High
Total	▓	▓	▓		
Identifying	▓	▓	▓	▓	▓
Using	▓	▓	▓		
Understanding	▓	▓	▓	▓	
Managing	▓	▓			

In Jerry's case, the results of the MSCEIT provided a confirmation and clarification of the issues involved in his leadership at that point in time. Jerry's scores on the identifying and understanding subscales were superb. That was no surprise: Jerry was excellent at perceiving how his staff felt about the move. He understood *why* his staff felt a loss, and how these feelings were changing (understanding emotion).

However, while Jerry was aware of, and understood, the issues, he was not integrating this emotional information into his decision making. He did not engage with these emotions, but instead blocked them out and relegated them to a lesser standing in his thought process. Jerry addressed each concrete problem, but not the *real*, underlying emotion-based problem: the sense that the team had been split apart, and cut off from Jerry.

The MSCEIT results, as well as the ability model, provided Jerry with both the insight and the process by which he could enhance his leadership style.

Using the MSCEIT for counselling and therapy

Advantages of using the MSCEIT for counselling and therapy

Assessment can assist the clinician to better pinpoint their client's strengths and weaknesses, and the MSCEIT provides an additional set of data in this process. As the MSCEIT uniquely assesses a client's emotional skills, it is especially suited for use in clinical settings.

Example of using the MSCEIT for counselling and therapy

Will, a 39-year-old attorney, was experiencing problems at home. His wife felt neglected and misunderstood by Will. At times, he could be smooth and sophisticated, whereas at other times, he was inappropriate and a bit 'off'.

Will spoke well. He was verbal, fluent and sophisticated. Will seemed to have a great deal of insight into himself and others. The therapist who was working with him found him delightful, and Will was brilliantly insightful in his therapy sessions. It was difficult to determine just what it was that gave Will so much difficulty in his life, as he appeared to be emotionally sophisticated and aware. It was at this point that Will took the MSCEIT. His MSCEIT scores were as follows.

MSCEIT score	Low				High
Total	▓	▓	▓		
Identifying	▓	▓			
Using	▓	▓	▓	▓	
Understanding	▓	▓	▓	▓	
Managing	▓	▓	▓	▓	▓

Will sometimes had a great deal of empathy for others (using emotions), and could feel what other people felt. His emotional vocabulary was superb, and if his beginning assumptions were correct, Will was able to accurately predict other's emotional reactions (understanding emotions). Will's decision making was usually right on target, but sometimes, seemed totally off the mark

(managing emotions). The success, or failure, of Will's judgments and decisions seem to stem from his initial 'read' of himself, or of other's, feelings. Often, this emotional read would be inaccurate, due either to Will's lack of awareness or attention, or to some other factor (identifying emotions).

Will's therapist now had identified the source of Will's interpersonal difficulties: they did not need to spend time on teaching Will emotion management strategies, nor did they have to help Will develop insight or empathy for others. Instead, the therapeutic work focused on creating greater initial awareness. Once Will was able to become more open to, and aware of, the emotional world, he was then taught how to accurately attend to others, to pick up on cues subtle and not so subtle, and to integrate this information to reach a conclusion about how other people were feeling.

Conclusion

The MSCEIT is 'different'. It defines EI differently, it measures EI differently, and it yields different sorts of results. For the practitioner, this means that great care needs to be taken when using the MSCEIT.

Note

Some of the material for this report was adapted from the book, *The Emotionally Intelligent Manager*, by David Caruso and Peter Salovey.

Developers, Users and Consumers Beware: *Warnings about the design and use of psycho-behavioural rating inventories and analyses of data derived from them*[1]

Kenneth J. Rowe

Australian Council for Educational Research

Dr Rowe is a Principal Research Fellow at ACER. Dr Rowe's post-graduate training in research methodology and design, educational psychology, assessment, measurement, psychometrics, and advanced statistical modelling, was undertaken at the University of London, and his doctoral research at the University of Melbourne. Dr Rowe's substantive research interests and expertise include: 'authentic' educational and psychological assessment; multilevel, 'value-added', educational/organisational performance indicators, achievement target-setting and benchmarking; teacher and school effectiveness; differential gender effects of schooling in the context of teaching and learning; the impact of externalising behaviour problems on students' learning outcomes in literacy and numeracy; and the educational/epidemiological implications of Attention-Deficit/ Hyperactivity Disorder (AD/HD) and Chronic Fatigue Syndrome (CFS) in children and adolescents.

Katherine S. Rowe

Department of General Paediatrics, Royal Children's Hospital, Melbourne

Dr Katherine Rowe is a Consultant Physician in the Department of General Paediatrics, and in the Centre for Adolescent Health at Melbourne's Royal Children's Hospital. In addition to her medical qualifications (MB.BS, MD), she has a Masters degree in Public Health (MPH), and a post-graduate Diploma in Child Development and Health and Welfare Education (DipEd) from the University of London, Institute of Education. Dr Rowe has also held academic appointments in the Department of Paediatrics at the University of Melbourne, consisting of teaching, clinical and research responsibilities, as well as the co-ordination of the undergraduate medical students' training program in paediatrics (1986-1997). Throughout these appointments, Kathy has developed extensive clinical and research experience in the management of children and adolescents with disabilities, behavioral and learning difficulties (including those with ADD and AD/HD), ear, nose and throat problems, as well as those with Chronic Fatigue Syndrome.

Abstract

Psycho-behavioural rating inventories are used routinely by psychologists and psychiatrists as assessment instruments to assist with the evaluation and 'diagnosis' of children and adolescents. They are also used in epidemiological studies to obtain normative/prevalence estimates of children/adolescents with psycho-behavioural 'problems'. Advantages entailed in their use include ease of administration and the convenience of obtaining estimates of normative behaviours from large numbers of informants. However, serious decisions are frequently made on the basis of 'measures' obtained from such instruments, including the labelling of a child as 'pathologic', subsequent referral to intervention therapy services, and prescription of medication by a physician. This workshop highlights key methodological issues endemic to the design and use of psycho-behavioural rating inventories, and the analyses of data derived from them. With specific reference to the assessment of **inattentive** behaviours, the workshop provides evidence indicating that traditional psychometric methodologies employed to construct 'scales' (typically from ordinal, item-response formats) and to report 'norms' that ignore the sampling, measurement, distributional and structural properties of the derived data, have long since passed their 'use-by-date'. Also demonstrated is that claims of *validity* and *reliability* employing these traditional methodologies can no longer be justified. Using data obtained from the administration of psycho-behavioural rating inventories in several large-scale research projects, these issues are illustrated and discussed in terms their substantive implications.

The outcomes of more robust methodologies are presented that stress the need to revise the design of child/adolescent psycho-behavioural rating inventories, and point to the adoption of more rigorous approaches to *measurement* and analyses of the related data.

1.0 Introductory comments

Psycho-behavioural rating inventories are used routinely by psychologists and psychiatrists as assessment tools to assist with the evaluation and 'diagnosis' of children and adolescents. They are also used in epidemiological studies to obtain normative/prevalence estimates of children and adolescents with psycho-behavioural 'problems', as well as for estimating effect magnitudes of the overlap between externalizing behaviour problems and educational under-achievement. Advantages entailed in their use include ease of administration and the convenience of obtaining estimates of normative behaviours from large numbers of informants. Nonetheless, serious decisions are often made on the basis of 'measures' obtained from such instruments, including the labelling of a child as 'pathologic', subsequent referral to intervention therapy services, and prescription of medication by a physician – all of which have potential impacts on students' cognitive, affective

and social/behavioural progress, especially in educational contexts.

Since behaviour at school and at home affects students' opportunities for learning and development, an enduring concern of teachers, parents and health professionals is the extent to which such maladaptive, externalising behaviours (particularly *inattentiveness*) adversely affect their learning outcomes. Students whose behaviours are regarded as inattentive, disruptive or maladjusted have been shown to be at risk of poor educational attainment.[2] Moreover, in addition to the consequences for an individual, behaviour problems in the classroom diminish educational opportunities for other students and contribute to teacher stress (Barkley & Pfiffner 1995a,b; Hinshaw & Nigg 1994; Brenner, Sörbom & Wallius 1985). Thus, in the context of clinical practice, as well as in psychosocial, epidemiological and educational, research, the measurement of child/student behaviour is of crucial importance.

The measurement of behaviour, however, is problematic. While it is possible to observe and estimate the frequency and saliency of specific behaviours by direct, objective means (see Rowley 1976, 1989), such approaches typically ignore the context in which behaviour takes place and fail to account for the possibility that some behaviours may be appropriate in certain circumstances and at certain stages of socio-behavioural development but inappropriate in others. Systematic observation techniques, particularly in school settings, are time-consuming and not practical options for screening large numbers of students.

In practice, child/student behaviour is assessed most frequently by means of rating inventories or 'checklists' completed by teachers, parents or clinicians (see Figures 2a, 2b and Figure 4). Typically, these multiple item inventories require response ratings in: (a) dichotomous categories (e.g., 'present'/'absent'; coded: '1' and '0', respectively); and/or (b) in Likert-type, ordered, polytomous categories of monotonically-increasing salience or frequency – coded: 0, 1, 2, 3, 4, etc (Likert 1932). Among the best known and widely used inventories in psychosocial, educational and epidemiological research include: the *Achenbach System of Empirically Based Assessment-ASEBA* (Achenbach & Rescorla 2001),[3] *Conners' Rating Scales* (Conners 1969, 1973, 1978, 1990a,b, 1994), *The Children's Attention and Adjustment Survey* (Lambert, Hartsough & Sandoval 1990), the *Rutter B(2) Scale* (Rutter 1967), the *Behaviour Problems Management System* (Galvin & Singleton 1984), and the *Rowe Behavioral Rating Inventories-RBRI* (Rowe & Rowe 1997a,b, 1999).

Although ratings on such instruments are essentially subjective, in the case of teachers and parents, this subjectivity is an asset since raters make 'in-context' judgments about behaviour against normative expectations and experience. Moreover, in addition to their convenience, behavioural rating scales for use by parents and teachers are indispensable, since child behaviours '...are almost always manifest in natural settings such as home and school, but might not be evident in laboratory or clinical environs. Parents' and teachers' judgments regarding the frequency, severity and appropriateness of children's behaviour are therefore essential for accurate detection and diagnosis...' (Edelbrock & Rancurello 1985, p. 429). The importance of teachers' roles in identifying, describing and defining child/student behaviour has long ago been expressed by Bower (1970, p. 94) as follows:

> The myth still exists that someone, somewhere, somehow knows how to assess behavior and/or mental health as positive or negative, good or bad, healthy or non-healthy, independently of the school context in which the individual is living and functioning. I strongly suspect that teachers, by focusing on the child's observable behavior in school, are closer to an operational reality of mental health than one can come up with in a sedentary examination.

2.0 Design problems endemic to typical behavioural rating inventories

Design problems endemic to typical behavioural rating inventories are at least twofold. First, for large-scale educational and epidemiological studies, a key disadvantage is their length. For obvious logistic reasons, inventories of thirty or more items with multiple response categories take considerable time to complete (e.g., Achenbach's CBCL/6-18 has 121 major items, and a further 26 'context/background' items). Completion of such inventories by teachers for all students in a class, for example, can be an arduous task and increase the likelihood of inaccuracies. Moreover, for longitudinal studies designed to investigate change in behaviour over time, it is necessary to use an inventory that is applicable to a wide age range. Inventories that have been designed to identify behaviours for specific age groups are not suitable for such purposes.

---

[2] For comprehensive reviews of this literature, see: Cantwell and Baker (1991); Elkins and Izard (1992); Hinshaw (1992a,b, 1994); Rowe (1991); Rowe and Rowe (1992a,b,b 1999); Singh, Ollendick and Singh (2002).

[3] The ASEBA comprises: the *Child Behavior Checklist for Ages 6 to 18* (CBCL/6-18), *Youth Self-Report* (YSR) and the *Teacher's Report Form* (TRF).

Second, a major disadvantage of most existing behavioural rating inventories is the use of items that focus exclusively on maladaptive rather than adaptive behaviours (e.g., Achenbach & Rescorla 2001; Conners 1969, 1973, 1994; Quay & Peterson 1975; Rutter 1967). Two examples are given in Figures 2a and 2b.[4] On the one hand this is not surprising given that such instruments are mostly constructed from the 'pathologic' (or negative) nomenclature contained in published manuals of diagnostic criteria for mental and behavioural disorders such as *DSM-II, DSM-III, DSM-III-R, DSM-IV* (APA 1968, 1980, 1987, 1994) and *ICD-9, ICD-10* (WHO 1978, 1992, 1996). In pointing to limitations entailed in the exclusive use of negatively-anchored items typical of most behavioural rating instruments, we have argued elsewhere:

> Emphasis on negative nomenclature is at the expense of a more balanced assessment and increases the risk of prejudicial searches for 'pathology', regardless of its presence or absence (Rowe & Rowe 1992a, p. 350).

Nor are such instruments independent of socio-cultural differences (Yao, Solanto & Wender 1988). For example, in a normative study of Achenbach's CBCL/6-18, Hensley (1988) found a consistent tendency by Australian parents to rate their child's behaviour as 'problematic' – significantly more so than their North American counterparts. Similar findings have been reported in comparative and normative studies of parent and teacher ratings (e.g., Glow 1978; Goyette, Conners & Ulrich 1978; Rowe & Rowe 1993a, 1997c; Verhulst & Akkerhuis 1989).

Item Nos. and Description		Response categories and coding			
Item No.	Item description	Not at all	Just a little	Pretty much	Very much
1	Restless and overactive	0	1	2	3
2	Excitable, impulsive	0	1	2	3
4	Fails to finish things he/she starts	0	1	2	3
5	Constantly fidgeting	0	1	2	3
6	Inattentive, easily distracted	0	1	2	3

Figure 2a Items from the *Inattentive/Overactive* sub-scale of Conners 10-item *Abbreviated Parent–Teacher Questionnaire* – ATPQ (n = 6923; α = 0.840)

Item Nos. and Description		Response categories and coding		
Item No.	Item description	Not true	Somewhat or sometimes true	Very true or often true
4	Fails to finish things he/she starts	0	1	2
8	Can't concentrate, can't pay attention for long	0	1	2
10	Can't sit still, restless, or hyperactive	0	1	2
41	Impulsive or acts without thinking	0	1	2
78	Inattentive or easily distracted	0	1	2

Figure 2b Items from the *Inattention and Hyperactivity-Impulsivity* sub-scale of Achenbach's *Child Behavior Checklist* – CBCL/6–18 (n = 6923; α = 0.777)

Apart from the negatively anchored wording, an interesting feature of the Conners' and Achenbach 5-item scales given in Figures 2a and 2b is the similarity of the constituent item nomenclature. However, the dissimilarity in the response formats – from a 4-category response (Conners' ATPQ) to a 3-category response (Achenbach's CBCL/6-18) – has had a notable effect on reducing the 'reliability' estimate (i.e., from α = 0.840 to = 0.777, respectively).[5]

More than 26 years ago Sandoval (1977) criticised the use of rating scales exclusively employing negatively worded items on the grounds that they are highly susceptible to rater bias and response sets such as 'reverse halo effects' or 'reverse generosity errors'. In a comparative study of format effects in rating scales of 'hyperactivity', Sandoval (1981) subsequently demonstrated that for positively worded items, raters are more willing to use the extreme rating categories for a given item, thus increasing the dispersion and discrimination of the ratings. In contrast, an inspection of the marginal distributions for negatively worded items show that they tend to be

---

[4]The data from which Cronbach's (1951) α 'reliability' coefficients for these scales have been computed derive from studies reported by Rowe and Rowe (1993c, 1995, 1997c, 1999).

[5]For comparative purposes, but with some reservations, the conventional estimate of 'internal consistency', namely, Cronbach's (1951) *standardised item alpha* (α), is given here. There are two major problems with the use of α: (1) the magnitude of α is a direct function of the number of items in a scale, regardless of their individual and shared error variance, and (2) α estimates of 'reliability' are lower-bound estimates, based on negatively-biased and inappropriate Pearson product-moment correlations among the constituent items – the data from which consist of responses in ordinal categories (see discussion in #3.0 below). For detailed expositions of the limitations of Cronbach's alpha in such circumstances, see McDonald (1981), Miller (1995) and Raykov (1997, 1998). For example, McDonald shows that: 'Proposals to regard coefficient alpha as a coefficient measuring homogeneity, internal consistency, or generalisability, do not appear to be well founded' (1981, p. 100). Similarly, Miller demonstrates '...the failure of α to meet certain basic criteria as an index of test homogeneity' (1995, p. 255).

highly skewed and leptokurtic – as illustrated in Figure 2c below.

Q1	Frequency	Percentage	Bar Chart
0	3875	56.0	□□
1	2325	33.6	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
2	587	8.5	□□□□□□□
3	136	2.0	□□

Q2	Frequency	Percentage	Bar Chart
0	2989	43.2	□□
1	2941	42.5	□□□
2	803	11.6	□□□□□□□□□□
3	190	2.7	□□□

Q4	Frequency	Percentage	Bar Chart
0	3968	57.3	□□□
1	2193	31.7	□□□□□□□□□□□□□□□□□□□□□□□□□□□□
2	528	7.6	□□□□□□
3	234	3.4	□□□

Q5	Frequency	Percentage	Bar Chart
0	4392	63.4	□□
1	1740	25.1	□□□□□□□□□□□□□□□□□□□□□
2	536	7.7	□□□□□□
3	255	3.7	□□□

Q6	Frequency	Percentage	Bar Chart
0	3607	52.1	□□
1	2519	36.4	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
2	584	8.4	□□□□□□□
3	213	3.1	□□□

Figure 2c. Univariate distributions for the 5-item *Inattentive/Overactive* sub-scale of Conners' 10-item *Abbreviated Parent-Teacher Questionnaire* – ATPQ: Parent ratings for 6923 children aged 5–16 years

Methodological and data-analytic problems

3.1 Factor analytic (FA) pathologies '

By far the most popular methodological means of defining and 'measuring' emotional and behavioural domains are via *exploratory* factor-analytic (FA) approaches to 'determine' the underlying dimensionality of multiple-item rating inventories administered by parents, teachers or clinicians. From Hinshaw's (1987) comprehensive review of 60 FA studies published between 1970 and 1986, it is interesting to note that all used *exploratory* FA approaches, and that 56 (93%) used orthogonal methods of factor extraction and rotation (mostly principal components analysis or principal factor solutions – both with varimax rotation). Such approaches are problematic on at least three grounds.

First, in the case of exploratory (unrestricted) methods of FA, the solutions are arbitrary, data-driven, hypothesis-generating, and invariably result in theory conflation (Jöreskog 1981; Rowe 1989, 2004). Second, orthogonal methods of factor extraction and rotation assume that the derived factors are uncorrelated or independent – by definition (Harman 1976). Since all items are allowed to load on more than one factor, the resulting correlated error variance alone is sufficient to yield shared variance across factors. Although the construction of uni-dimensional scales is highly desirable from a measurement perspective, given the considerable literature concerning the non-independent and overlapping dimensions of child behaviour and psychopathology, orthogonal FA methods are difficult to justify either

substantively or empirically. At best, such procedures yield discrepant findings that are all-too-frequently ignored or interpreted as 'statistical artifact'. At worst, such procedures yield mis-specified and misleading estimates that contain large proportions of measurement error, with crucial implications for substantive interpretations of findings from subsequent statistical modelling.

Third, and perhaps most serious of all, such methods are invariably applied to item responses in dichotomous or 3 to 5-point Likert-type ordinal categories, and rely on the computation of Pearson product-moment (PP-M) inter-item correlation matrices – estimated by default in most omnibus statistical packages. What is overlooked in such instances is that the assumptions underlying PPM correlations (i.e., normal distribution and homogeneity of variance) are always violated (see: Jöreskog 1994; Rowe 2002, 2004a; Rowe & Rowe 1992a, 1997c, 1999). Indeed, failure to take account of the measurement and distributional properties of response variables in factor analysis, amounts to what Hendrickson and Jones (1987) refer to as 'an undisciplined romp through a correlation matrix' (p. 105). Consistent with the insights of Scarr (1985), we have suggested elsewhere: 'Given the almost universal application of these procedures, it could be argued that current claims to substantive knowledge about dimensions of child psychopathology may be little more than the products of methodological and statistical artifact' (Rowe & Rowe 1992a, p. 351). Whereas there is evidence for awareness of this problem among some researchers in child psychology and psychiatry, it is rare, and warnings about

such violations have remained patently unheeded. For example, Morris, Bergan and Fulginity (1991, pp. 373-374) attempted to alert their fellow researchers in the following terms:

> Traditional factor analytic procedures assume that manifest indicators are normally distributed continuous variables. Test items are generally dichotomous or polytomous variables that reflect no more than an ordinal scale. Thus, a normal distribution cannot be assumed. Traditional practice has been to ignore the requirement of continuous normally distributed variables and to factor analyze test items. The result of this approach is biased estimates of model parameters.

A number of approaches are now available that provide ways to carry out confirmatory factor analyses with ordinal data and obtain unbiased estimates of model parameters. Applications of these techniques with clinical assessment instruments are largely lacking. Thus, the state of affairs that exists at present is that little attempt has been made to establish the construct validity of large numbers of clinical assessment instruments that are used with children. … Of particular concern is the issue of the validity of using existing assessment instruments for referral, diagnosis, treatment selection, forensic evaluations, and the evaluation of treatment outcome.

Further, from Jöreskog (1994, p. 383), the special features of ordinal variables are worth noting:

> Observations on an ordinal variable are assumed to represent responses to a set of ordered categories, such as a five-category Likert scale. It is only assumed that

a person who responds in one category has more of a characteristic than a person who responds in a lower category. *Ordinal variables are not continuous variables and should not be treated as if they are. Ordinal variables do not have origins or units of measurement. Means, variances, and covariances of ordinal variables have no meaning* (our emphasis).

It is common practice to treat scores 1, 2, 3, 4, representing the ordered categories of an ordinal variable as numbers on an interval scale and use a covariance matrix computed in the usual way to estimate a structural equation model. What is so bad with this is not so much that the distribution is non-normal; more importantly the distribution is not continuous: there are only four distinct values in the distribution. The use ordinal variables in structural equation models (SEM) requires other techniques than those which are used for continuous variables.

It should also be noted that, in general, SEM techniques (including both exploratory and confirmatory factor analysis) assume that the observed data are quantitative variables measured, at least approximately, on an interval scale, and whose distributions are approximately multi-normal. In most psychosocial research applications, however, the observed variables are typically non-normal and/or of mixed response types: categorical, ordinal (Likert-type ratings) and continuous. Under such circumstances, the use of ordinary product-moment correlations is not appropriate (Brown 1989; Healy & Goldstein 1976). Instead, *tetrachoric* (dichotomous with dichotomous) *polychoric* (ordinal with ordinal)[6] and *polyserial* correlations (ordinal with continuous) should be computed, and

---

[6]Unlike the product-moment correlation which is a measure of association (or standardised co-variation) between the 'scores' for two continuous variables, the polychoric correlation is an estimate of joint variation '..in the latent bivariate normal distribution representing the two ordinal variables' (Jöreskog & Sörbom, 1988, pp. 1–9). For further technical details related to the estimation of polychoric correlations, see Jöreskog (1994), Olsson (1979), Poon and Lee (1987).

the correct asymptotic covariance matrix of such correlations should be analyzed by the method of Weighted Least Squares (WLS), using PRELIS (Jöreskog & Sörbom 2003a), for example. Failure to do otherwise can lead to gross errors in correlation estimates, distorted parameter estimates, and incorrect goodness-of-fit measures and standard errors (Huba & Harlow 1987; Jöreskog & Sörbom 2003b).

Hence, when the data on observed items/indicators are non-normal and non-continuous (e.g., dichotomous, ordinal/polytomous categories), the use of product-moment correlations is inappropriate (Jöreskog 1990, 1994), yielding large negative biases in their estimates (Carroll 1961; Jöreskog & Sörbom 1979, 1988; Lord & Novick 1968). An illustration of the negative bias entailed by the use of PP-M correlation estimates compared with their polychoric counterparts is given in Tables 3a and 3b – using the five items from the *Inattentive/Overactive* sub-scale of Conners' 10-item *Abbreviated Parent–Teacher Questionnaire* (ATPQ) given in Figure 2a. In this case, compared with the polychoric correlations, the PPM correlations are negatively biased by 0.1 (on average).

Table 3a Lower Triangular Matrix of PPM Inter-correlations Among Conners' Inatten/OA Items

Items	Q1	Q3	Q4	Q5	Q6
Q1	1				
Q2	0.621	1			
Q4	0.408	0.360	1		
Q5	0.597	0.484	0.481	1	
Q6	0.466	0.415	0.659	0.546	1

Table 3b Lower Triangular Matrix of Polychoric Inter-correlations Among Conners' Inatten/OA Items

Items	Q1	Q2	Q4	Q5	Q6
Q1	1				
Q2	0.734	1			
Q4	0.497	0.424	1		
Q5	0.697	0.585	0.575	1	
Q6	0.563	0.492	0.769	0.656	1

In brief, as a consequence of the typical inappropriate use of PP-M correlation estimates for dichotomous or ordinal variables, instead of their consistently less biased *tetrachoric* or *polychoric* counterparts, respectively, substantial negative bias (i.e., under-estimates) in the inter-item correlations and subsequent factor parameters is unwittingly introduced.

These moribund approaches, that have long-since passed their 'use-by-date', lead to at least two major problems when modelling relationships among composite scale scores, or to compare the magnitudes of their interdependent effects. First, the unit-weight addition of indicator variables in the formation of the scale scores ignores the possibility that indicators typically contribute differentially to the measurement of composite/scale 'scores'. Second, the unit-weight addition of indicators may invalidate the composite score if one or more of the indicators 'measure' a construct other than the one under consideration. Behavioural rating developers and researchers who continue to use 'data-fishing' methods that fail to account for the measurement, distributional and structural properties of the obtained data (typically consisting of raw, un-weighted response scores on Likert-type item/indicators), run the risk of generating biased and misleading estimates (Hendrickson & Jones 1987; Morris, Bergan & Fulginity 1991; Rowe 2002, 2004a; Rowe & Rowe 1992a,b, 1997c, 1999; Table 3a).

During the past 25 years, these problems have been minimised somewhat by the use of confirmatory factor analysis (see Bentler 1980; Bollen 1989; Jöreskog 1981, 1990; McDonald 1978, 1985; Muthén 1989). The advantages of confirmatory factor analysis (CFA) methods over exploratory factor analysis (EFA)

approaches for such purposes are well documented and need not be reiterated here, but for relevant discussions, see Bollen (1989), Gorsuch (1983), Marsh (1987, 1994), Marsh and Grayson (1994), Rowe (1989, 2002, 2004a), Rowe and Rowe (1992a, 1997, 1999), Scott Long (1983), and Stevens (1995). In brief, the advantages include: '...the ability to formulate, define specifically, and test an a *priori* model; the ability to selectively specify or estimate particular model parameters; and the opportunity to directly test and compare the relative goodness of fit of competing models' (Stevens 1995, p. 217). CFA models allow for unequal contributions of indicators towards the measurement of latent variables (e.g., *inattentiveness*) and the models will fit only when the indicator variables associated with any one latent variable are valid indicators of that latent variable. Further, when the number of indicator variables becomes large, parameter estimation and model fit statistics are unstable unless the sample size is also large.

3.2 Scale 'score' 'pathologies'

A further problem in applied research relates to the widespread use of scale 'scores' derived from behavioural rating inventories for the purposes of classification and diagnosis. Typically, scale 'scores' are computed as *factor scores* (from factor analysis), or worse, as simple, unit-weighted, additive indices (or counts) of their indicators, regardless of either the measurement or distributional properties of the constituent indicators, or their relative contribution to the scale 'score'. Illustrations of the distributional characteristics of unit-weighted scale 'scores' from two behavioural rating inventories are provided in Figures 3a and 3b next page.
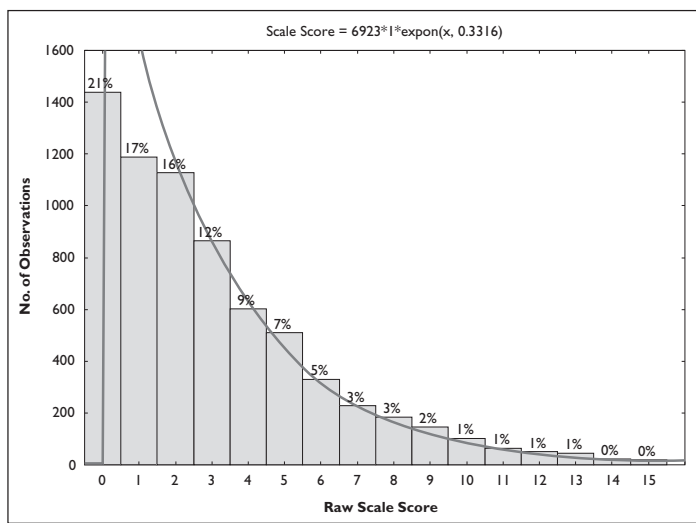
Figure 3a. Distribution of raw scale 'scores' from the five *Inatten/OA* scale items from Conners' *Abbreviated Parent-Teacher Questionnaire* – ATPQ: Parent ratings for 6923 children aged 5-16 years (Min-Max: 0-15)
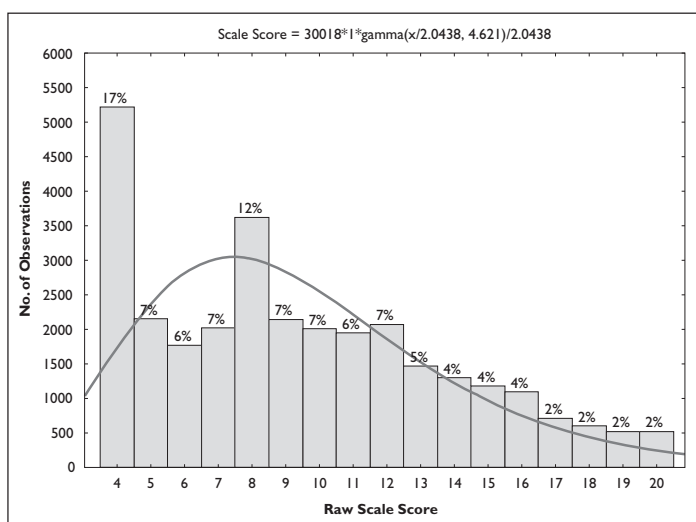

Figure 3b. Distribution of raw scale scores from the four *Attentive-Inattentive* scale items of the *RBRI 12-Item Teacher Form:* Teacher ratings for 30,018 children – aged 5-16 years (Min-Max: 4–20)

From Figures 3a and 3b, it is clear that the distributions of the raw scale 'scores' are non-Normal. That is, the score distribution for Conners' *Inatten/OA* scale (Figure 3a) indicates that the 'best-fit' to the data is described by a negative exponential function, whereas the distribution for the RBRI *Attentive-Inattentive* scale scores (Figure 3b) is best described by a gamma function.[7] All too frequently, these 'scores' are then treated (inappropriately and incorrectly) as Normally-distributed continuous variables in omnibus applications of the general linear model, which further assume that both the constituent item indicators and the computed scale

'scores' are 'measured' without error (Rowe 1989). In such cases, it is well established that the use of standard Normal deviate estimates to describe the distribution of scale 'scores' is misleading (see: Johnson, Kotz & Balarkrishann 1994, 1995; Kendall & Stuart 1963).

Due to the inherent complexity of behavioural disorders in childhood, Ullmann *et al.* (1985) have argued that the common use of a single 'cutoff' score on a rating scale to diagnose deviance is inappropriate and misleading [e.g., a score of 15 on Conners' ATPQ to 'diagnose' *Attention-Deficit/Hyperactivity Disorder* (AD/HD). Although it is customary to select two standard deviations from the mean for these purposes, such selections are arbitrary and can be modified depending on whether one wishes to minimise false positives or false negatives. This approach has been aptly illustrated by Szatmari, Offord and Boyle (1989) in their review of eleven studies reporting prevalence rates of AD/HD. Four of these studies employed diagnostic 'cutoff' scores of 1.0, 1.5 or 2.0 standard deviations from the mean, in the absence of substantive criteria for doing so, 'resulting in the identification of different numbers and types of cases' (Szatmari *et al.* 1989, p. 221). For example, reported prevalence rates for AD/HD vary from less than 1% (Rutter, Tizard & Whitmore 1970), 14.3% (Trites, Dugas, Lynch & Ferguson 1979), to as high as 20% (Shaywitz & Shaywitz 1991), depending on: (1) the methods of data collection, (2) the sampling characteristics of the populations targeted, and (3) the arbitrary determination of deviance criteria.

Further, variability in measurement and 'cutoff' scores, together with sampling differences, lead to substantial

---

[7]Despite the problems associated with computing simple additive 'scale scores' discussed here, the advantages of employing bi-polar item nomenclature formats (as used in the RBRIs) is evident from Figure 3b – especially in terms of discrimination. Note that the special design features of the RBRI are outlined in #4.0.

differences in prevalence estimates. In the context of predictive or explanatory research, there is little rational justification for identifying, *a priori*, a fixed proportion of the child population as 'AD/HD', for example, particularly when such a dimension is more meaningfully viewed as a continuum, both in quantitative and qualitative terms. Despite the utility and obvious convenience of rating scales, especially for large-scale survey research, the psychometric limitations endemic to their common design, construction and use seem to be largely unrecognised by most developers, users and researchers.

In sum, 'cut-off' scores based on commonly used statistical criteria (i.e., + 1 ≤ SD ≤ + 2) are arbitrary since they are dependent on the properties of the 'measures' used, as well as on sampling variability across studies. Such arbitrariness leads to substantial differences in prevalence estimates that may or may not reflect actual problems. Moreover, given that all behavioural 'measures' computed in this way are highly skewed (as illustrated in Figure 3), statistical criteria of these kind are difficult to justify due to the inevitable violation of the assumptions of *normality* of distribution and *homogeneity* of variance.

3.3 Measurement and scale construction problems

It is important to stress that the foundation of ALL responsible data analysis and statistical modelling is ***good measurement*** and the ***minimization of measurement error variance***—otherwise, what is generated are serious 'garbage-in' 'garbage-out' problems that unjustifiably conflate theory and yield misestimated parameters (at best) and misleading 'findings' (at worst). This is especially the case for analyses of data obtained from behavioural rating inventories (Rowe & Rowe 1992a, 1997c, 1999), as well as for agencies and/or health professionals wishing to use data to identify performance indicators of 'health' or 'pathology', particularly for intervention and policy purposes (see: Rowe 2001, 2004b; Rowe & Lievesley 2002). It should also be noted that measurement error problems are **seriously compounded** with 'contextual' or 'compositional' variables that are aggregated from the characteristics of level-1 units (i, e.g., students) within level-2 units (j, e.g., classes or schools), because the measurement error inherent in the level-1 variables is averaged across the level 1 units in each level-2 unit, or higher (see Rowe 2004b). Moreover, there is additional sampling error whenever $nj < Ni$ – which is always the case.[8]

It is now well established that factor-analytic (FA) and Classical Test Theory (CTT) approaches to measurement and scale construction in psychosocial inquiry do not even meet the three **basic 'requirements'** of ***measurement***, namely: (1) the need to focus on only ***one way*** in which objects or persons differ in terms of an attribute of interest; (2) the need for a ***unit of measurement*** (so that equal numerical differences represent equal amounts); and (3) ***objectivity*** (freedom from the characteristics of the instrument and of the person(s) undertaking the measurement).[9] Further, it has been demonstrated that FA and CTT approaches are **not** commensurate with modern measurement theory and practice (see especially: Embretson 1996; Embretson & Hershberger 1999; Masters & Keeves 1999; Wilson & Engelhard 2000; Wright 1999). Key reasons for this are beyond the scope of this paper. Nevertheless, in brief, scale score meaning via CTT and FA approaches is merely inferred from norm-referenced 'standards'. That is, the scores *per se* have no meaning for what an assessed individual does or can do; moreover, such scores are *sample-dependent*.

By contrast, from item-response approaches to measurement (better known as *Item Response Theory* models – IRT), the scale scores are *sample-independent* and score meaning can be referenced directly to the constituent items – from which a linear scale can be constructed and described qualitatively (e.g., Masters 2001a,b; Masters, Meiers & Rowe 2003; Stephanou 2000). Following the seminal work of Thorndike (1904), Thurstone (1926) and Guttman (1944), the 'requirements' of objective measurement in the psychosocial sciences have been promulgated by the Danish mathematician Georg Rasch (1960), who laid the foundations of what has become known as *modern measurement theory*, or *Rasch measurement*. The advantages of this approach to measurement are noted in more detail later in #5.0.

4.0 Improving the design of psycho-behavioural rating inventories

Due mainly due to the poor design, low reliability and lack of predictive validity

---

[8] Note that Fuller (1987) provides a comprehensive account of methods for dealing with measurement errors in linear models, and Goldstein (1995, chp. 10) extends some of those procedures to the multilevel modeling case.

[9] For comprehensive treatments and applications of modern measurement theory (including Rasch measurement), see: Embretson and Hershberger (1999), Masters (1982, 1988), Masters and Keeves (1999), Masters and Wright (1997), Rasch (1960, 1977), Stephanou (2000, 2002), Wilson & Engelhard (2000), Wright (1999), Wright and Mok (2000). For excellent introductory overviews, see Masters (2001a,b). For an application to psycho-behavioral rating inventories, see Smith and Johnson (2000).

of the Conners' and Achenbach's negatively-worded types (as illustrated earlier in Figures 2a and 2b), the *Rowe Behavioral Rating Inventories* (RBRIs)[10] were developed from empirical research applications to obtain valid and reliable 'in-context' measures of child/student *externalising* behaviours for use in clinical settings, as well as in educational, psycho-behavioural and epidemiological research. Since the rationale for the development and use of the RBRIs has been comprehensively documented and demonstrated by Rowe (1991, 1997a) and by Rowe and Rowe (1992b,c, 1993c, 1994b, 1995, 1997c, 1999), the need for reiteration here is not required. However, for illustrative purposes, Figure 4 records the constituent 4 items of the *Attentive-inattentive* scale from the RBRI *12-item Teacher Form*.

Three features of the design and item content of the *RBRI*s should be noted. First, following the semantic bipolar format advocated and used by Kysel, Varlaam, Stoll and Sammons (1983), the RBRI items allow for assessments both adaptive and maladaptive behaviours (i.e., *health* and *pathology*). Second, the item nomenclature has been formulated on the basis of extensive cross-validations of parent and teacher descriptions of typical child/student externalizing behaviours at home and at school, and in three domains: *sociable-irritable/antisocial, attentive-inattentive*, and *settled-restless*. Third, the items are applicable to a wide age range, having been developed from comprehensive trialing and application among large samples of children/students (n > 180,000) in the age range of 5 to 16 years.

How the *RBRI* forms should be scored depends on the purposes for which they are to be used, but a major advantage of the bipolar item format is that alternative methods for item scoring may be used.[11] That is, in studies concerned with the measurement of *maladaptive* behaviours, Item Nos. 2, 7 and 10 shown in Figure 1c may be scored 1 to 5 (from left to right) on the five-point ordinal scale, with scoring reversed for Item No. 1. In such cases, a low score on each item reflects positive adjustment and a high score, poor adjustment. In studies concerned with the effects of *adaptive* behaviours, the items may be scored such that high scores are reflective of positive adjustment.

5.0 Improving the measurement properties of behavioural rating inventories

At this point, a brief discussion of the utility of fitting behavioural rating data to item-response measurement models that meet the basic requirements of *objective measurement* is helpful. In particular, what is highlighted here is the utility of *Rasch measurement* in constructing scales by calibrating item indicators with dichotomous and/or polytomous response categories – typical of behavioural rating inventories. For relevant work in this area, see references cited in footnote 9.

Teachers, for each of the following paired behavioral statements, please mark a cross over the dot (e.g., ⊗) which is **nearest** the statement that **best** describes the TYPICAL behavior of THIS student at school

1. Cannot concentrate on any particular task; easily distracted	O O O O O	Can concentrate on any task; not easily distracted
2. Perseveres in the face of difficult or challenging tasks	O O O O O	Lacks perseverance; is impatient with difficult or challenging tasks
7. Persistent, sustained attention span	O O O O O	Easily frustrated; short attention span
10. Purposeful activity	O O O O O	Aimless; impulsive activity

Figure 4. *Attentive-inattentive* items from the *RBRI 12-item Teacher Form* (n = 30,018; α = 0.926)

---

[10] The RBRI inventories consist of two major rating forms: (1) a 16-item Teacher Form, and (2) a 20-item Parent Form, for use in clinical settings with children in the age range of 5 to 16 years. Both forms are supported by an accompanying interactive computer software package, RBRI Profile® (Rowe & Rowe, 1997a) and a User's Manual (Rowe & Rowe, 1997b). Similar information is provided for two shorter versions of these forms, namely the 12-item Teacher Form and the 16-item Parent Form – devised specifically for use in the large-scale monitoring and epidemiological research.

To date, the research applications include epidemiological studies of the relationship between the ingestion of synthetic food dyes and behavioral change in pediatric populations (Rowe KS, 1988, 1996; Rowe KS & Briggs, 1992, 1993; Rowe & Rowe, 1994a,b), and in studies of factors affecting student literacy and numeracy achievement (Crévola & Hill, 1998a; Hill et al. 1993, 1996; Hill & Rowe, 1996, 1998; Rowe, 1991, 1997; Rowe, Fullarton et al., 2003; Rowe & Hill, 1998; Rowe & Rowe 1992b,c, 1993, 1995, 1997c, 1999). In these studies, the inventories have been validated for dye-challenge and for monitoring the comorbidity of externalizing behaviors and academic under-achievement.

The psychometric and normative properties of the *RBRI*s are based on cross-validated and replicated samples of teacher ratings for 33,433 school-aged children in five age cohorts (5–6, 7–8, 9–11, 12–13, 14–16 years) and parent ratings on 16,569 children across the same age cohorts. Data on concurrent parent and teacher ratings have been obtained for 9566 children. Specific details of the samples, data properties and related research applications are available in the RBRI User's Manual (Rowe & Rowe, 1997b) and in Rowe and Rowe (1999).

[11] A further advantage of employing a bipolar item format is that it minimises the occurrence of 'negative halo effects' by minimising the risk of prejudicial searches for 'pathology'.

The work of Rasch and those who have followed has impacted radically on the *theory* of *measurement*, and especially on applications in educational and psychological assessment (psychometrics). In brief, the Rasch approach to the *measurement* of a *latent* or *composite* variable – derived from responses to multiple items/indicators in dichotomous or polytomous categories – is that it allows for *scale construction* by calibrating jointly the location of each item **and** respondent on an empirical scale of increasing attribute (e.g. *performance, extroversion, attentiveness, attitude*, etc.). Fitting indicator-response data to Rasch's logistic model yields an unbounded **logit scale**[12] (with interval properties) that allows any pair of items (and person pairs) to be compared in terms of the magnitude of the interval difference between their locations on the scale. An illustration of this feature is the 'Person-item map' provided in Figure 5 [print-out from *ACER-QUEST*, Adams and Khoo (1999)] that not only facilitates the setting of 'cut-scores', or 'pass marks' on assessments, but also, 'benchmarks' and/or performance standards, for example.

A particular advantage of Rasch-calibrated scales is that empirical, evidence-based evaluations can be made of the extent to which each item or indicator contributes to the measurement of the latent variable being constructed (i.e., *differential item/indicator functioning* in terms of measurement accuracy). A further advantage is that a scale so constructed allows detailed **descriptions** of performance levels or standards to be made in both quantitative and qualitative terms (e.g., Masters 2001a,b; Masters, Meiers & Rowe 2003; Stephanou 2000, 2002). The properties of Rasch-calibrated scales are

such that items from separate assessment sources/occasions of the same kind (e.g., performance standards) can be **equated** and located on a common measurement scale – provided

that some indicators and/or respondents (cases) overlap, or are linked from one assessment to another. These procedures are known as *common-item equating* and *common-case equating*, respectively.
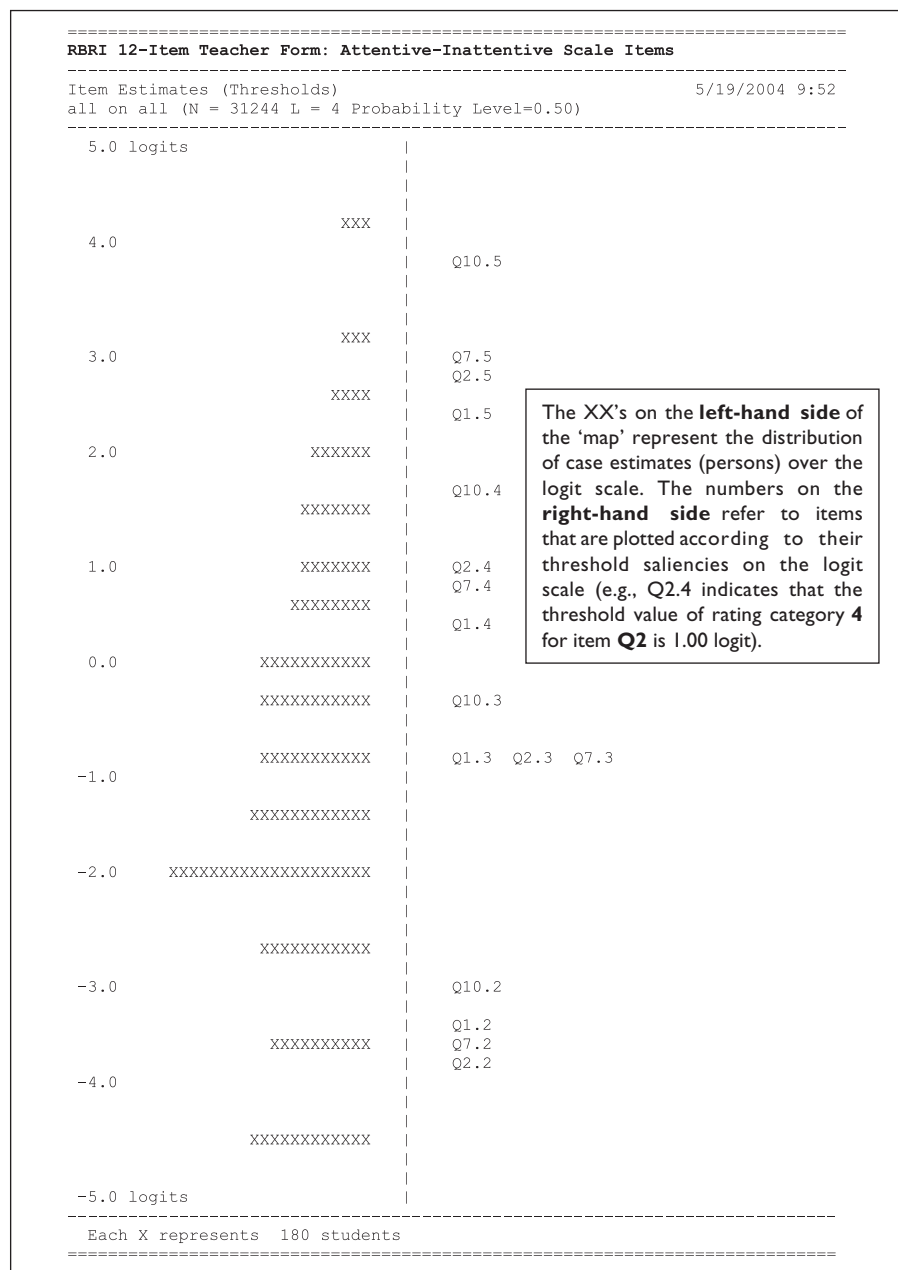
```
================================================================
RBRI 12-Item Teacher Form: Attentive-Inattentive Scale Items
----------------------------------------------------------------
Item Estimates (Thresholds)                        5/19/2004 9:52
all on all (N = 31244 L = 4 Probability Level=0.50)
----------------------------------------------------------------
    5.0 logits              |
                            |
                            |
                       XXX  |
    4.0                     |
                            |   Q10.5
                            |
                            |
                            |
                       XXX  |
    3.0                     |   Q7.5
                            |   Q2.5
                      XXXX  |
                            |   Q1.5
    2.0              XXXXXX  |
                            |
                            |   Q10.4
                     XXXXXXX |
                            |
    1.0              XXXXXX  |   Q2.4
                            |   Q7.4
                    XXXXXXX  |
                            |   Q1.4
    0.0           XXXXXXXXXXX|
                            |
                XXXXXXXXXXX  |   Q10.3
                            |
                            |
                XXXXXXXXXXX  |   Q1.3  Q2.3  Q7.3
   -1.0                     |
                            |
                XXXXXXXXXXX  |
                            |
   -2.0    XXXXXXXXXXXXXXXXX |
                            |
                            |
                XXXXXXXXXXX  |
   -3.0                     |   Q10.2
                            |
                            |   Q1.2
                XXXXXXXXXX   |   Q7.2
                            |   Q2.2
   -4.0                     |
                            |
                            |
                XXXXXXXXXXX  |
                            |
   -5.0 logits              |
----------------------------------------------------------------
    Each X represents  180 students
================================================================
```

> The XX's on the **left-hand side** of the 'map' represent the distribution of case estimates (persons) over the logit scale. The numbers on the **right-hand side** refer to items that are plotted according to their threshold saliencies on the logit scale (e.g., Q2.4 indicates that the threshold value of rating category **4** for item **Q2** is 1.00 logit).

Figure 5. 'Person–item map' of *Attentive–Inattentive* **scale items from the RBRI** *12-item Teacher Form* **for 30,018 children – aged 5–16 years**

[12]The *logit* is a unit of measurement derived from the natural logarithm of the odds of an event, where the odds of that event is defined as the ratio of the probability that the event will occur to the probability that the event will not occur. A logit scale is used in both educational and psychological assessment because it has interval scale properties. That is, if the 'difficulty' or 'salience' of an assessment item (e.g., Item A) is 1.0 logit greater than the difficulty or salience of Item B, then the odds of an individual responding correctly (or more saliently) to Item B are 2.7 times the odds of the same individual responding correctly (or more saliently) to Item A, regardless of whether this person has high or low ability/attribute. Similarly, if the ability or attribute of Person A is 1.0 logit greater than the ability of Person B, then the odds of Person A responding correctly (or more saliently) to an item are 2.7 times the odds of Person B responding correctly (or more saliently) to the same item, regardless of item difficulty or its salience.

These properties of scales constructed via *Rasch measurement* are **especially** useful in the development of item banks from which items and/or indicators of known attribute salience can be drawn to develop further assessment instruments that are comparable. It is also extremely valuable (and **vital**) for applications in: (1) longitudinal, repeated-measures studies of the same cases, and (2) cross-sectional studies of different respondent cohorts at different times. Such procedures are not possible using traditional Classical Test Theory (CTT) methods, and have considerable advantages over traditional methods based in CTT – particularly those employing factor analytic approaches. For these reasons, *Rasch measurement* is used as the basis for constructing and describing scales for all cognitive, affective and behavioural assessment instruments developed by the Australian Council for Educational Research (ACER), as key elements in its national and international work in assessment and reporting.[13]

6.0 Concluding comments

In highlighting key methodological problems endemic to the design and use of psycho-behavioural rating inventories, and analyses of data derived from them, the purpose of the present paper is twofold.

First, it is argued that the design features of most psycho-behavioural rating inventories used routinely by epidemiologists, psychiatrists and psychologists to assess children and

adolescents with psycho-behavioural 'problems' are less than adequate. In particular, the almost exclusive use of negative item nomenclature in such inventories increases the risk of prejudicial searches for 'pathology', regardless of its presence or absence. Given that serious decisions are frequently made on the basis of 'measures' obtained from such instruments, including: the labelling of a child as 'pathologic', subsequent referral to intervention therapy services, and prescription of medication by a physician, it is crucial that such instruments be of the highest quality in terms of both their design and measurement properties.

Second, on the basis of supporting evidence the paper argues that traditional *Classical Test Theory* and *factor-analytic* methodologies employed to – construct 'scales', 'measure' behaviour, report 'norms' and to specify 'cut-off' scores for the purposes of 'classification', 'diagnosis' and the provision of prevalence estimates – have long since passed their 'use-by-date'. Indeed, it is argued that claims of *validity* and *reliability* employing these traditional methodologies can no longer be justified. Rather, the need to adopt more rigorous approaches to *measurement* and analyses of the related data is urgent. It is hoped that both the traditional 'emperors' of psycho-behavioural inventory design, development and data-analytic methodology, and we, the product users, will heed such cries about our 'nakedness' before our sartorial delusions render our efforts ludicrous.

References

Note: due to space limitations here, the entire document - including the references cited in this paper - are available for download in pdf format from www.acer.edu.au/research/programs/learningprocess.html

---

[13]A full listing of psycho-behavioral and educational assessment instruments developed by ACER via *Rasch measurement*, are available at: www.acer.edu.au/tests/index.html. For examples of ACER's national and international work in assessment and reporting (among many others), see: Adams *et al.* (1988-1997); Lokan, Greenwood and Cresswell (2001); Masters (2002); Masters and Forster (1996, 1997); Rowe and Stephanou (2003). For examples of ACER's international publications and work programs with OECD, IEA and the World Bank, visit: www.acer.edu.au/about/international.html. Specific ACER projects and publications related to assessment and reporting are available at: www.acer.edu.au/research/reports.html, and www.acer.edu.au/research/programs/assessment.html.

Online Assessment: *Current Issues and Future Views*

Kevin Chandler

Chandler and McLeod

Kevin, a registered Psychologist since 1973, is the founding managing director of Chandler Macleod Group. He has extensive experience in the fields of organisational psychology and organisational change and has lectured extensively in the areas of personality and understanding individual differences including guest lectures in statistics at Macquarie University. He has conducted organisational assignments in large and small organisations resulting in significant productivity improvement and cost savings.

Kevin has been retained by organisations such as Accenture, Telstra, Australia Post, Leighton Holdings and others to advise on numerous organisational programs that have focused on leadership culture, change management, organisational strategy and organisational values.

He is a Member of the Australian Psychological Society, Fellow of the Australian Institute of Management and Fellow of the Institute of Management Consultants in Australia. He is a past Federal President and Victorian President of the Institute of Management Consultants.

Current Issues

For most of the employed population we forget that as little as five years ago the use of email and the Internet as a business tool was not all pervasive as it is today. The rapid expansion of the worldwide web and the development of ASP processes have changed forever the nature of business. The flow on into psychological assessment, whilst initially slow, has been accelerating (particularly over the last two years) with many of the world's major test providers now offering ASP services to their test users.

Issue	Verdict
Online versus pen & paper	a
Proctored versus unproctored testing	Jury is out
Jobs based test batteries and competency modeling	a
Online validation	Jury is out

Online versus pen and paper

Computers and the internet offer test users a system that overcomes some of the problems associated with pen and paper test administration. These problems can be identified as:

* distribution and security of test materials;
* training of test administrators;
* the use of timed tests when untimed tests, particularly of abilities and aptitudes, might be more suitable;
* the inability to run job specific test batteries where sizable numbers of candidates are being tested at the one time;
* scoring/errors;
* access to limited norming data;
* data storage; and
* difficulties in candidate follow up for research and validation purposes.

Psychologists over the last half of the last century (seems like only yesterday) grew up with these issues and after a while stopped thinking about the impact they had on the administration of assessment tools. We learnt to live with the frustrations. But some testing requirements were rigorously adhered to. There was consistent use of a standardised testing environment with test users understanding issues of fatigue, distractibility and the like. That includes running test batteries in the same order, starting them at the same time each day where possible, and generally trying to ensure that all candidates had the opportunity to do their best. This process also requiring proof of identity for candidates completing assessment.

By and large the careful application of psychological assessment together with close evaluation of background, skills, knowledge and experience does lead to better recruitment outcomes. During the 90s I conducted many research projects evaluating the success of standard recruitment procedures without using psychological assessment and identified that when asked the question 'If you knew at time of hire what you now know, would you re-hire?' 68 percent of the working population would be rehired. However, this left a substantial 32 percent where supervisors felt there was a mismatch. The application of psychological assessment tools in these research exercises lifted this success ratio to 93 percent.

So we can safely advise that the supervised application of psychological assessment based on the selection of job-related assessment tools results in substantial improvement in on job

performance – and at a cost that was substantially lower than the cost of obtaining the candidate (advertising, interviewing, agency fees and the like).

With the advent of the Internet and the ASP we were able to address many of the weaknesses in pen and paper administration.

- Test batteries can be tailored to make them job specific, even when there are a number of other test candidates in a test room.

- Test administration times come down – on average for a full test battery of three ability tests, an interest test and a personality test the test battery can be completed approximately 25 percent faster online, because there is less paper shuffling and candidates can be more focused rather than having to spend time waiting around for the slowest test candidate to be ready for each new test.

- The system can take over the role of test administration with the responsibilities of the test supervisor now vastly lessened. The system can ensure that candidates fully understand what is required of them by halting the progress on a particular test if the candidate has not adequately completed the practice examples. Often this was glossed over historically.

- Scoring and transposition errors are removed due to the 'no human intervention' in the scoring and data storage process.

- Service delivery has improved dramatically as there are now many more locations where a candidate can undertake assessment.

But what are the worrying trends?

The pervasiveness of the internet allows for the distribution of poorly validated and unreliable test instruments by unqualified assessors. As a delivery mode, the Internet creates a perception of professionalism and sophistication, which may not be matched when the theoretical constructs of the various test batteries are explored. Within Australia we are seeing the distribution of assessment tools by non-professionals who are not required to answer to the same ethical standards as registered psychologists or members of the Australian Psychological Society. This is no more evident than in the area of the provision of unsupervised testing of candidates for graduate or entry level jobs.

Proctored versus unproctored testing

The significant issue here is the devaluing of the contribution that can be made to the recruitment decision by online tools, particularly where their interpretation is made by a qualified psychologist. This kind of self-reporting for issues as serious as whether one individual gets a $35,000–$40,000 job over another is not countenanced in other parts of our society.

We don't allow athletes to bring their own samples for a drug test. We do not allow individuals to complete university exams sitting at home in their own time. We don't allow individuals to pronounce themselves fit to drive, we don't allow individuals to deliver financial services without licenses. In fact, in many parts of our lives where our behaviour can interact negatively on others or where distortion of our capabilities can result in a negative imposition on others we expect regulation and regimentation. The worrying trend towards the administration of psychological assessment tools in an unregulated, unsupervised environment is a real cause for concern amongst professionals. Try taking a glucose tolerance test at home

and sending the results off to your doctor. It just doesn't happen.

This isn't the case with this year's crop of undergraduates. Many are being sent questionnaires at home and asked to complete psychological assessment as a first cut form of screening. Apart from the fact that we have shown time and again that intelligence equates to perhaps 25 percent of the variance in on job performance and that other techniques should be used for screening down candidates, the opportunities for error are substantial. The errors can be mitigated against if the shortlist is to undergo further evaluation. However, when candidates realise that recruiters are using this data to screen them out their behaviour will follow the reward system, and the system will be compromised as it has been in the area of drug testing in sport.

Our results in trialling this form of testing are concerning. Last year this kind of assessment for graduate candidates for a major bank led to more than 20 percent of candidates complaining to us about issues relating to the equipment or the system that was being used. Would you be concerned about the fairness or otherwise of this system particularly if you realised these results were being used as a screen-in screen-out technique?

At the moment it is questionable whether candidates see the test session as nothing more sinister than an extension of their application form and some of the open-ended questions they may be asked to complete there. Do they see it as anything different from the magazine questions relating to 'How good a lover am I?' It is not seen to be cheating to look up the answers and see that answering B all the way through will get you a high score, whereas answering A or C won't. We cannot blame the candidates if we promote a system that has inherent weaknesses.

Job-based test batteries and competency modelling

Another key area addressed by Online Assessment relates to one of the inherent weaknesses in the way computerised testing operated during the 90s. The interpretation of test scores in the absence of an evaluation of the work environment does not facilitate effective decision-making. The Internet has fostered the development of online competency modelling to assist the psychologist in identifying cultural and environmental factors and then inputting this information into the decision making model. Research has identified a number of critical competencies that differentiate high versus low performers in jobs. These competencies tend to be job rather than industry or function specific. Online competency modelling is much more efficient in terms of its application and enables us then to tailor both the test batteries according to the competency models as well as tailoring reporting around the competency model. This introduces a dimension to the assessment process, which has been difficult to apply manually.

Online validation

A final area where online assessment can leap ahead over traditional methods is in the area of validation. The difficulties in securing client cooperation in obtaining performance data has been a major impediment to the establishment of ongoing validation of manually administered test batteries. The latest online assessment systems incorporate regular follow up of candidates following assessment. This normally happens at three months and 12 months with a view to identifying the suitability or otherwise of the candidate in the role. On this basis, modifications to test batteries can be made quickly and efficiently and clients can gain an even greater appreciation of the value and contribution made by psychological assessment in improving the overall productivity of their workforce.

In summary, the paradigm shift we all hope to achieve with online assessment has only just started.

Online Assessment: *Current Issues and Future Views*

Aimee Williamson

Saville Holdsworth Ltd (SHL)

Aimee Williamson is a registered psychologist who holds a Masters in Organisational Psychology. Aimee is currently the Product Business Director with SHL, a leading provider of psychometric assessment products designed for selection and development of people at all levels in organisations. Aimee has worked with SHL for four years, and is responsible for managing the SHL product portfolio in Australia. Prior to her current role, Aimee was responsible for the portfolio of internet based products offered by SHL. Prior to joining SHL, Aimee spent two years working for UBS, a global investment bank, in Human Resources and four years with the Department of Veterans' Affairs.

Context

The internet offers an attractive administration alternative for many types of assessment, including testing for schools, universities and testing in the workplace. This paper focuses on the use of online assessment in the workplace. More specifically, the focus of this paper is on the use of online assessment for high-stakes decision making in the workplace, such as selection decisions. Although the internet offers many applications for assessment in the workplace beyond selection decisions (for example, 360 degree feedback, career counselling and development interventions), it is in the area of selection testing that key challenges and issues surrounding online assessment are experienced.

Naglieri, Drasgow, Schmidt, Handler, Prifitera, Margolis and Velasquez (2004) highlight the difference between testing and assessment. Whereas testing can be defined as the administration, scoring and interpretation of results, assessment is the integration of a range of information about an individual in order to develop an overall recommendation, or judgement. Naglieri et al. (2004) further propose that although the internet is well developed for the purpose of testing, it is not yet sufficiently evolved to deal with complex psychological assessment. For the purposes of this paper, the emphasis will be placed on online testing.

Introduction

The internet has been firmly established as a viable delivery platform for a range of psychological tests. Organisations are attracted to the core benefits the internet is said to offer, those being speed, cost and convenience in the testing process. Research has shown that the number of organisations who are looking to introduce the use of the internet in their testing programs is

likely to double over the next three years (SHL, 2003). Research has also shown that there are risks associated with the use of the internet as a testing delivery platform. These risks, whilst well researched by the psychological community, have not always been effectively communicated to the organisations relying on the results of the tests. This paper explores a number of the risks associated with the use of the internet for testing purposes. It explores options for reconciling the convenience and speed of the internet with the fundamental tenets of validity and reliability. Finally, some views of future developments for internet testing will be presented.

Online testing: key issues

> Current psychometric standards, particularly those regarding test reliability and validity, apply even though the way in which the tests are developed might be different (p. 2, Naglieri et al., 2004).

The internet is a delivery mechanism for psychometric tests. It is an alternative to paper and pencil and computer based testing. Web enabled tests should reflect the same psychometric standards of reliability and validity as tests that are delivered through other mediums. The concern to maintain the application of psychometric standards is at the heart of each of the issues presented below.

Equivalence of online tests

Establishing the equivalence of computer-based tests with those administered via paper and pencil has been widely researched. In summary, research supports the notion that equivalence can be demonstrated for the majority of tests when administered

in a paper and pencil and computerised environment. Bartram (2004) reports that non-cognitive tests, and cognitive tests that are not highly dependent on speed, have been shown by numerous studies to be equivalent. Equivalence of computer-based forms of tests has been a more significant issue for speeded tests, where the nature of the task may change dependent on the administration mode (Bartram, 2004). In these tests, it is important to consider the test design and norming in order to accurately measure the intended ability.

Although there has been significant research conducted in relation to the equivalence of computer and paper and pencil based tests, there is a paucity of research dealing with issues of online equivalence. The research that is available (although limited) suggests that where appropriate care has been taken in the development of the tests online, there is unlikely to be any major impact on the equivalence of results on the basis of internet enabled delivery (Bartram, 2004).

In conclusion, there is a need for additional research on the equivalence of psychometric instruments when placed on the internet. However, research to date suggests that internet tests are likely to be equivalent to those administered in a paper and pencil environment.

Testing in an unproctored environment

Perhaps the key factor that attracts many organisations to the use of psychometric testing online is the potential to reach candidates in geographically remote areas. They are seeking unproctored test administration. Paradoxically, unproctored test administration is the area that causes the most significant concern from psychologists. The key issues associated with unproctored test administration can be summarised as follows:

- Who is taking the test?
- What environment is the individual taking the test in?
- Are candidates collaborating and sharing the test's intellectual property?
- Is it possible for candidates to tamper with the internet platform to adjust settings (e.g. time)?

The challenge for psychologists is to explore ways to reconcile the psychometric challenges posed by unproctored testing with the desire of organisations to conduct it. Evidence collected through both research and practice suggests it may be possible to reconcile these positions to some degree. (This is discussed in Reconciling Risks and Opportunities – see below.)

The diminishing emphasis on reliability and validity

The internet has reduced barriers to entry for new players within the test development and test publishing marketplace. An increasingly diverse marketplace offers organisations a number of advantages through increased choice, and more competitive pricing. However, this change in the market has also led to a dilution of the fundamental principles of reliability and validity. There are a range of tests offered online that are not supported by appropriate reliability and validity evidence.

This issue for psychologists is one of continuing education. How can psychologists communicate the fundamental importance of reliability and validity to businesses and organisations that are focused on efficiency and cost? The challenge is to translate reliability and validity into tangible results for organisations and to demonstrate, often in dollar terms, the returns that organisations can gain by utilising those tests that reflect sound psychometric principles.

Reconciling risks and opportunities

As the number of tests available online increases, and organisations recognise the potential benefits of internet testing, the demands on psychologists to provide robust, yet innovative, solutions will increase. The issue of unproctored assessment is one area in which psychologists have worked from a research and practice perspective, to identify ways in which to harness the potential of the internet, whilst minimising risks and maintaining psychometric principles. The following discussion explores work that has been done in this area.

Research on the impact of unproctored personality assessment

Bartram and Brown (2003) conducted a study to investigate the impact of unproctored administration using the OPQ32i (an ipsative response format questionnaire). Bartram and Brown (2003) compared matched samples of candidates who completed the OPQ32i in proctored and unproctored situations. Their results showed that there was no impact on scale means, reliabilities or scale intercorrelations when the OPQ32i was administered in an unproctored environment.

This research implies that the presence or absence of a proctor is unlikely to influence the response patterns to ipsative questionnaires to a significant degree. These results have formed the basis of a policy allowing the administration of the OPQ32i under unproctored, but controlled, conditions over the internet. In all cases, it is recommended that a validation interview be conducted with the candidate.

The use of Item Response Theory and item banks to protect intellectual property

The use of IRT provides a possible solution to the question: *Are candidates collaborating and sharing the test's intellectual property?*

- IRT offers a number of benefits over classical test theory as a method to construct tests. In relation to selection testing, there are two key opportunities offered by the use of IRT.
- IRT enables the development of more efficient tests, providing employers with shorter tests yet reliable results; and
- IRT can be used to develop tests that are reusable, with secure item content (Bartram, 2004).

SHL has used IRT to develop online, randomised ability tests that can be used for screening purposes. The nature of the construction of these tests means that each individual will receive an equivalent, but different test variant. Tests constructed in this manner are not subject to the issue of compromised test content. As each candidate receives a unique variant of the test, it is not possible for candidates to collaborate and share the test items. The implication of this increased security of test items is that a test such as this, developed using IRT, is more suitable than a classical test for online administration in an unproctored environment.

Using appropriately constructed unproctored ability tests as a screening measure

Traditional methods for screening candidates in recruitment processes include resume screening, and academic reference checking. Both of these approaches are subject to significant

falsification by candidates, and in some cases the information provided is difficult to verify. Furthermore, the validity of these approaches for predicting performance on the job is low.

Cognitive ability tests have been shown to be the single best predictor for performance on the job (Robertson & Smith, 2001). For large-scale recruitment, such as graduate recruitment, where job relevant abilities are critical to job performance, the potential to sift on the most valid predictors early in the recruitment process is an attractive one.

SHL (2002) conducted a study of a large-scale graduate employer in Australia. Over 1200 candidates completed two online ability tests in an unproctored environment as an initial screen. Candidates who reached a pre-determined benchmark on the unproctored tests were subsequently re-tested in a proctored environment, using a different ability test. The results of 741 candidates who completed both the unproctored screening tests and the proctored ability tests were correlated. The correlations obtained between the tests are presented below.

	Sifting numerical	**Sifting verbal**
Supervised verbal	.24**	.47**
Supervised numerical	.62**	.22**

**P<0.01

These results indicate a high level of construct validity for the unproctored tests. Furthermore, examination of the pattern of responses on both tests did not indicate a significant proportion of false positives on the screening tests. This suggests that there was not a high

incidence of cheating in the unproctored environment.

At present, there is no way to verify that the individual who takes a test in an unproctored environment is the intended candidate. However, the use of an appropriately constructed online ability test to screen candidates, followed by verification testing in a proctored environment, can provide employers with the benefit of early and accurate screening, along with confirmation of the ability levels of their selected applicants.

Conclusion

New methods made possible by emerging technologies will push the boundaries of existing psychometric theory and it is up to psychologists to test and expand the limits of psychometrics to keep pace with these innovations (p. 2, Naglieri et al., 2004).

The introduction of the internet offers a wide range of benefits to organisations seeking to use psychometric testing as part of their selection process. Offsetting these benefits, there are numerous challenges to address to ensure that psychometric standards are maintained whilst the opportunities of the internet are maximised. Innovation will be the key to maximising the power of the internet in the future. The use of Item Response Theory is becoming more prevalent online, and over time it is likely that the use of adaptive testing, particularly for personality questionnaires will increase. As suggested by Naglieri et al. (2004), it is essential that psychologists test the boundaries of psychometrics to ensure that the opportunities offered by the internet are fully harnessed.

References

Bartram, D. (2004). Computer based testing and the internet, draft paper.

Bartram, D. & Brown, A. (2003) Online testing: Mode of administration and the stability of the OPQ32i. BPS Occupational Psychology Conference, Bournemouth, Compendium of Abstracts, p.63.
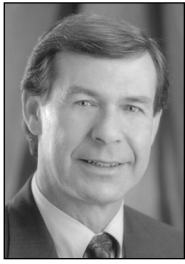
Naglieri, J.A., Drasgow, F., Schmidt, M., Handler, L., Prifitera, A., Margolis, A. and Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist, 59*, 150–162.

Robertson, I.T. and Smith, M. (2001). Personnel Selection, Journal of Occupational and *Organisational Psychology, 74*, 441–472.

SHL (2003). Internal market research.

SHL (2002). Unpublished validation study.

Uses and Advanced Interpretation of the Stanford-Binet Intelligence Scales, Fifth Edition (SB5)

Gale H. Roid

Vanderbilt University

Gale Roid is the Dunn Professor of Educational and Psychological Assessment, Department of Special Education, Peabody College of Vanderbilt University, Nashville Tennessee. Author of many well-respected publications including *Stanford-Binet Intelligence Scales Fifth Edition*, Professor Roid has also co-authored the *Stoelting Brief Nonverbal Intelligence Test (S-BIT)* and the *Leiter International Performance Scale –Revised.*
Simpson College & Graduate School, USA
Author of SB5

New editions of nationally standardised tests provide modern wording, illustrations, enhanced measurement procedures, updated theory and research, and new standardisations, enhancing the validity of test interpretations. After a seven-year revision project, the Stanford-Binet Intelligence Scales – Fifth Edition (SB5) (Roid 2003a) was published with enhanced features, norms, and procedures. This workshop provides an introduction to the design, theory, uses, and clinical interpretation of the SB5 by the author of the Fifth Edition.

The Stanford-Binet (Terman 1916), and its Binet predecessor (Binet & Simon 1908), have become international classics in the history of test development. Until the 1970's when many competing instruments became available, the Stanford-Binet was the most widely used intelligence scale worldwide. The SB5 continues the tradition of Binet and Stanford-Binet editions. The new edition combines the point scale format of the Fourth Edition (SB4) by Thorndike, Hagen, and Sattler (1986) with the age-level format found in previous editions such as the classic Form L-M (Terman & Merrill 1937, 1960). Examiners begin a standard test administration by giving the two routing subtests: Object-Series/Matrices (nonverbal) and Vocabulary (verbal). Estimates of ability in the nonverbal and verbal domains are obtained from raw scores on each routing test and used to tailor the remaining assessment to the examinee's functional ability. Simple conversion tables show the examiner which functional levels (ranging from easy to hard, levels 1 through 6) of the nonverbal and verbal scales to continue testing. By adding the nonverbal routing test and an entire one-half of the SB5 in the nonverbal domain, the new edition provides excellent features for testing individuals with limited English or communication difficulties. The nonverbal section requires a 'low language demand', that is minimal receptive language and mostly nonverbal responses by the examinee (pointing, moving pieces, etc.). Also, the two-stage testing procedure, with routing subtests first and functional levels sections next, provides highly precise estimates of cognitive ability in a relatively short period of time.

Thus, the SB5 is a wide-ranging, individually administered test battery. Norms were designed for ages 2 through 85+ years and the subtests cover five cognitive factors – fluid reasoning, knowledge (crystallised ability), quantitative reasoning, visual-spatial ability, and working memory – in both the verbal and nonverbal domains. Importantly, the SB5 is the first intellectual battery to cover five cognitive factors in both the nonverbal and verbal domains. Five nonverbal subtests and five verbal subtests measure each of the factors.

Many new features have been added to SB5 and features of previous editions enhanced. Many of the new features were designed to enhance the usefulness of the SB5 for assessments with preschool children, those with mental retardation, as well as individuals with intellectual giftedness.

For example, colorful toys were added and the number of blocks and manipulatives increased for both children and low-functioning individuals. The upper end of each subtest was extended with challenging items in fluid reasoning, vocabulary, quantitative, and memory areas. In addition, many features were added to make the test easy for examiners to administer and score such as record form designs similar to other intelligence batteries. An extensive

interpretive manual (Roid 2003b) was added, as well as a book in the Wiley 'Essentials of Assessment' series (Roid & Barram 2004). Another introductory chapter will be published this year (Roid & Pomplun 2004). Training materials for universities are also available from Riverside Publishing.

Psychometric properties

Extensive studies of reliability, validity, and fairness were conducted as part of the SB5 standardisation. The main technical features of SB5 are briefly outlined here and presented in detail in Roid (2003c). The normative sample for SB5 included 4,800 subjects, ages 2 to 96 years. The highest age grouping employed in the norm tables was 85+. The composition of the normative sample closely approximated the stratification percentages reported by the United States Census Bureau (2001). Stratification variables included gender, geographic region, ethnicity (African-, Asian- and Anglo/Caucasian-American, Hispanic, Native American, and Other), and socioeconomic level (years of education completed or parent's educational attainment). Additionally, subjects were tested (N = 1,365) from officially documented special groups such as individuals with mental retardation, learning disabilities, attention deficit, and speech or hearing impairments.

Internal consistency reliability ranged from .95 to .98 for IQ scores and from .90 to .92 for the five Factor Index scores. For the 10 subtests, average reliabilities (across age groups) ranged from .84 to .89, providing a strong basis for profile interpretation. Split-half reliability formulas were used for subtests and composite reliabilities for IQ and Factor scores. Test-retest and inter-examiner reliability studies were also conducted and showed the stability and consistency of SB5 scoring. Nonverbal IQ show smaller practice effects (differences between pre- and

post-test scores) than other intelligence batteries, allowing more frequent retesting (e.g., every 6 months rather than 1-year intervals).

Evidence for content-, criterion-, and construct-related validity of SB5 was detailed in Roid (2003c) including extensive studies of concurrent, predictive, and factorial validity. Also, good evidence of consequential validity and fairness of predicting achievement was were reported in Roid (2003b). Examples of validity include the correlations with other assessment batteries such as .90 with SB4, .90 with a 5-factor version of the WJ III general cognitive score (Woodcock, McGrew, & Mather 2001). The correlations with Wechsler scales are also quite substantial and range from .83 to .84– similar in magnitude to the concurrent correlations observed for other major intelligence batteries. The substantial predictive correlations (median .75) between SB5 and two major achievement batteries (WJ® III and WIAT® -II) provide a strong basis for comparing intellectual and achievement scores of individuals.

Extensive studies of the factor structure of SB5 were conducted, including confirmatory factor analyses using LISREL 8.3 (Joreskog & Sorbom 1999). The factor analyses were calculated for five successive age groups (2–5, 6–10, 11–16, 17–50, and 51+) comparing factor models with 1, 2, 3, 4 and 5 factors. The five factor models showed superior fit including the non-normed fit (NNFI) index (ranging from .89 to .93), comparative fit index (CFI) ranging from .91 to .93, and root mean square error of approximation (RMSEA) ranging from .076 to .088.

Administration and scoring

The workshop will cover details of test administration using the SB5 Record

Form. The author will demonstrate administration of key subtests. The standard order of administration is reflected in the record form layout, beginning with the two routing subtests and continuing into the 'levels' portion of the SB5 (first the nonverbal and then the verbal functional levels). The levels are organised into subscales called 'testlets' – four at each level – representing the cognitive factors measured by SB5. The subscales are added together to form the subtest raw scores. Raw scores are converted to Wechsler-style scaled scores (mean 10, SD 3) for use in the 10-subtest profile of examinee results. Scaled scores are then summed to derive the IQ and Factor Index scores.

The seven-step interpretive strategy

The workshop will cover interpretation and present several interesting case studies. Most interpretative methods for intelligence scales begin with the global (full scale) scores. However, due to the strength of the nonverbal sections of the SB5, we suggest using a seven-step method that first emphasises the differences between nonverbal and verbal scores. Each of the seven steps is briefly described below with examples of their application and a case study to emphasise the key features of interpretation.

Step 1: assumptions

The first assumption of professional assessment is that standardised instructions have been followed exactly. When examiners change standardised procedures, the use of normative interpretations are at risk. The assumption of a valid and standardised assessment must be examined by all users before proceeding with interpretation. As outlined in Braden and Elliott (2003), accommodations

may be necessary (and their description is outlined in the SB5 Interpretive Manual, Roid 2003b).

Step 2: background and context

The ethnic, gender, religious, cultural, or other characteristics of the examinee's background may greatly affect test interpretation. To fully understand the implications of this step, a case study will be used to highlight important considerations. Suppose you have been asked to give the SB5 to a high-school student we will call Noor. She is 16 years, 10 months old and is a native of Pakistan who immigrated to the United States two years ago with her parents who were both college educated. Although she spoke English during the testing session, her native language is Urdu and she is enrolled in an English-as-second-language program in her high school.

Examinees with unique cultural backgrounds or recent immigration such as Noor have a variety of levels of acculturation into Western culture. Acculturation is the process of adapting to a new culture and involves various levels of acquiring or rejecting behaviours and attitudes of the host culture (Dana 1993). Therefore, examiners should be aware that acculturation status must be assessed and use interview techniques or other methods of discovering what levels and types of acculturation have been acquired by the examinee.

Step 3: nonverbal IQ vs. verbal IQ

NVIQ should be compared to VIQ using the typical methods of examining the statistical significance of the difference and its rarity in the normative population (Sattler 2002). Extensive tables for evaluating such score differences among IQ, Factor Index and subtest scores of SB5 were presented in the SB5 Technical Manual (Roid 2003c), and included in the SB5

scoring software. Consider the case study of Noor that was introduced in Step 2. Her Nonverbal IQ was 97, 19 points higher than her Verbal IQ of 78, a difference that is statistically significant and relatively rare in the normative sample. Any report of Noor's results should highlight the statistical significance and clinical importance of her 19 point difference (NVIQ versus VIQ) and mention her language and cultural context. Instead of placing more emphasis on the Full Scale IQ (which was 87 for Noor), the Nonverbal IQ of 97 probably represents a lower bound of her cognitive potential because of the English-language difficulties she has.

Step 4: full scale IQ

The FSIQ provides the most global, summary index of general cognitive ability across the five cognitive factors measured by SB5. It is also the most reliable index of all SB5 scores because it is based on all parts of the test and research shows it to be internally consistent (technically, at the .98 level on a scale from .00 to 1.00). Investigators of intelligence theory, such as Carroll (1993) would say that FSIQ estimates the 'g' (general ability) that is found underlying all the scores within typical cognitive test batteries. Also, experts caution that general ability, reflected in FSIQ, can be affected by many environmental factors that can either increase or decrease cognitive performance. For example, poverty or cultural deprivation, illness or accidental injury, violence or abuse can decrease the cognitive functioning in an individual whereas wealth, health, and a protected environment can promote cognitive growth. For these and other reasons, FSIQ should never be presented to the examinee, parents, guardians, teachers, or others as a static, life-time, unchanging quality represented by a single number. Instead, all IQ scores

should be presented within a range of possible scores (the confidence interval).

Step 5: factor index scores

Use the five factor index scores at the next level of interpretation for several important reasons. First, the factor index scores are more reliable than individual subtest scores (in the .90 to .92 as compared to .84 to .89). Second, the factor index scores are based on extensive research in cognitive abilities (e.g., Carroll 1993; Horn & Cattell 1966) spanning nearly 50 years. Third, the metric of these factor index scores is the typical standard score with mean 100 and standard deviation 15 used on many different tests; hence ease of comparison across multiple tests. Fourth, the cognitive factors measured by the SB5 align with those of the CHC theory (Flanagan 2000) and the cross-battery approach developed by McGrew and Flanagan (e.g., McGrew & Flanagan 1998). In the cross-battery approach, individual subtests (or, combinations of subtests that measure a factor) can be administered from two or more cognitive test batteries and the results for a particular individual combined when making interpretations.

Step 6: subtest comparisons

Roid (2003c) used the methods of Aldenderfer and Blashfield (1984) to derive sets of profile patterns from the normative sample and showed that differentiated profile patterns can be found in 40% to 50% of individual profiles in large samples. The key to finding these differentiated profiles was to employ more sensitive cluster analysis methods using correlations as similarity indexes. This allowed interesting profiles to emerge more clearly, e.g., profiles high in verbal versus nonverbal subtests.

Another reason for clinicians to look for patterns among the subtest scores of

SB5 is the relatively high level of reliability of the scores compared to other cognitive batteries. The average internal consistency reliabilities of the SB5 subtests range from .84 to .89, which compares favorably to other batteries where subtests may have reliabilities in the .70 to .80 range.

Step 7: qualitative interpretation

Clinicians are very creative in developing ways of interpreting sequences of item responses, problem solving styles, and various behavioural reactions to the testing situation. These creative explorations are part of a qualitative interpretation of SB5 as contrasted to a score-based interpretation. The three main strategies suggested for qualitative interpretation of SB5 are (1) using test-session behaviour to temper interpretations of test scores, (2) 'testing the limits' by re-testing or interview procedures following the completion of standardised administration of SB5, and (3) various interpretations specific to certain subtests.

Advanced subtest interpretation

Several advanced methods of interpretation are used on the subtest profile scores of intelligence batteries and will be briefly reviewed in the workshop. These methods include (a) the relationship between each subtest and the underlying general-ability factor ('g'), (b) the specificity (specific variance) of each subtest which helps to evaluate and identify the subtests that have unique qualities not shared by the other subtests, and (c) use of composite scores formed from combinations of subtests, e.g., to predict early emergence of learning disabilities. Formulas and examples of the latter will be presented.

Summary

Thus, the SB5 workshop will prepare the attendees for administration, scoring, interpretation, and general use of the new edition. Theory and research on applications of SB5, including case studies of interesting clinical cases will be included. Resource materials will be distributed as handouts. An overview of the computer scoring software, the test manuals, record form, and other ancillary materials will be presented in summary.

References

Aldenderfer, M. S., & Blashfield, R.K. (1984). *Cluster Analysis*. Newbury Park, CA: Sage

Baddeley, A. (1986). *Working Memory*. New York: Oxford University Press.

Binet, A., & Simon, T. (1908). Le development de l'intelligence chez les enfants. *L'Année psychologique 14*, 1–94. (Also published in A. Binet, & T. Simon, (1916). *The Development of Intelligence in Children 14*, 1–94. (E.S. Kite, Trans.). Baltimore: Williams & Wilkins.

Braden, J.P., & Elliott, S.N. (2003). Accommodations on the Stanford-Binet Intelligence Scale, Fifth Edition. (Assessment Service Bulletin No. 2). Itasca, IL (USA): Riverside Publishing.

Carroll, J.B. (1993). *Human Cognitive Abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Dana, R.H. (1993). *Multicultural Assessment Perspectives for Professional Psychology*. Boston: Allyn & Bacon.

Embretson, S.E. (1996). The new rules of measurement. *Psychological Assessment 8*, 341–349.

Flanagan, D.P. (2000). Wechsler-based CHC cross-battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler test scores. *School Psychology Quarterly 15*, 295–329.

Horn, J.L., & Cattell, R.B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology 57*, 253–270.

Joreskog, K.G., & Sorbom, D. (1993). LISREL 8: User's reference guide. Chicago, IL: Scientific Software International.

McGrew, K.S., & Flanagan, D.P. (1998). *The Intelligence Test Desk Reference (ITDR): Gf-Gc cross-battery assessment*. Boston: Allyn & Bacon.

Roid, G.H. (2003a). *Stanford-Binet Intelligence Scales, Fifth Edition: Examiner's manual*. Itasca, IL: Riverside Publishing.

Roid, G.H. (2003b). *Stanford-Binet Intelligence Scales, Fifth Edition: Interpretive manual*. Itasca, IL: Riverside Publishing.

Roid, G.H. (2003c). *Stanford-Binet Intelligence Scales, Fifth Edition: Technical manual.* Itasca, IL: Riverside Publishing.

Roid, G.H., & Barram, R.A. (2004, in press). *Essentials of Stanford-Binet (SB5) Assessment*. New York: Wiley.

Roid, G.H., & Pomplun, M. (2004, in press). Interpretive strategies for the Stanford-Binet 5th Edition. In D.P. Flanagan, J.L. Genshaft, & P. L. Harrison (eds), *Contemporary Intellectual Assessment* (2nd Ed.), New York: Guilford.

Sattler, J. M. (2002). *Assessment of Children* (4th ed.). San Diego, CA: Author.

Terman, L.M. (1916). *The Measurement of Intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Scale*. Boston: Houghton Mifflin.

Terman, L.M., & Merrill, M.A. (1937). *Measuring Intelligence*. Boston: Houghton Mifflin.

Terman, L.M., & Merrill, M.A. (1960). *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M*. Boston: Houghton Mifflin.

Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). *The Stanford-Binet Intelligence Scale: Fourth Edition*. Itasca, IL: Riverside Publishing.

United States Census Bureau. (2001). Census 2000 summary File 1 United States. Washington, DC: Author.

Woodcock, R.W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

Issues in Conceptualising and Assessing Emotional Intelligence

Jennifer Bryce

Australian Council for Educational Research

Dr Jennifer Bryce is a Research Fellow at ACER where she has worked in the area of assessment and reporting since 1992. She also has professional interests in teaching and learning practices and has recently completed projects involving case study work in areas such as lifelong learning, arts education and mental health. Jennifer is currently working with Doug McCurry on a project for the Victorian Curriculum Assessment Authority that involves conceptualising and assessing generic skills. In 2003 she completed a PhD entitled *Constructing Intra and Interpersonal Competencies in a Context of Lifelong Learning.*

Introduction

The author is in the process of developing a construct of the 'emotional intelligence' needed by novice workers. Driving this research is a belief that emotional intelligence is valued in today's workplace particularly because of the need for flexibility and adaptability in terms of career prospects and because today knowledge is so readily accessible and constantly changing that young people need to have strong generic skills that will help them to keep learning throughout their lives.

The intent is to develop a robust construct that will form a basis for rigorous high stakes assessment. It will be argued that some existing measures of emotional intelligence are not entirely suitable for use at the point of transition to the workplace from secondary school or university. They were not designed for this purpose, thus many of these tests are formative/diagnostic and self-report, for example, Bar-On (1997). Various existing measures of emotional intelligence and personal skills will be discussed in this light.

A robust construct of emotional intelligence needs to be made up of conceptions for which there is a shared understanding of meaning and, being an 'intelligence', the conceptions need to be defined as kinds of cognition rather than personality attributes. Thus, as Mayer et al. (1999) have argued, such conceptions need to:

- be capable of being operationalised as a set of abilities;
- have components that can be intercorrelated and be related to pre-existing intelligences (while also showing some unique variance); and
- have abilities that develop with age and experience.

Being forms of cognition, such conceptions should be able to be learned, or at least enhanced, in schools.

As a first step towards developing the construct, personal skills have been socially constructed. This process will be outlined. The paper will then go on to argue how and why such personal skills should be conceptualised as a form of cognition. The paper will conclude with a demonstration of the kinds of items that can test these areas.

A desired outcome from this research would be the incorporation of assessment of 'emotional intelligence' at the end of secondary school.

Social construction of personal skills: a first step

The author will describe fieldwork undertaken to socially construct the personal skills needed by novice workers. This is seen as a first step to conceptualising 'emotional intelligence'. Interviews and focus group discussions were conducted with young people who had started work straight after school and also young people who had started professional work after completing a university degree. In addition, human resources managers and 'professional representatives' involved with recruitment were interviewed.

The major question discussed was:

- What personal skills are needed when young people enter the workforce in today's environment of rapidly changing knowledge?

Supporting questions were:

- What do employers look for when recruiting school leavers and university graduates?
- What intra /interpersonal skills do young people believe they need when they start work?

The broad categories of personal skills that emerged from these discussions were Communication, Working with Others (interpersonal) and Self-Management and Self Confidence (intra personal). These categories were expected, as they have emerged from other studies where employers have been asked similar questions, for example Wellington (1994), A.C. Nielsen (2000). The purpose for this study was to probe the meaning of these conceptions and to amplify the definitions of these concepts with facet descriptors (McCurry & Bryce, 1997). The outcomes will be presented and discussed.

What is new about this social construction of personal skills?

The broad categories that emerged from discussions with novice workers and recruiters are similar to those outlined in much of the literature reporting on surveys of employers and the deliberations of education bodies, such as the Ministerial Council on Education, Employment, Training and Youth Affairs and the National Goals of Schooling in Australia (Adelaide Declaration 1999). But discussions probed the meaning of these broad terms and the analysis defined the conceptions by using facet descriptors. The purpose was to look at a particular conception from several different angles – in what ways might a person excel at this personal skill? There have been many surveys of employers, asking them what skills they look for when recruiting, the outcomes of some of these will be referred to, in particular the frequently quoted response: 'hire the smile and the attitude and we will train the rest'. The research tried to find out exactly what is meant here by 'the smile' and 'the attitude'.

A novel aspect of this research was the inclusion of novice workers. These recently employed young people could readily recall the 'steep learning curves' and other significant experiences that occurred during their first few months of employment. Some facets of personal skills were mentioned by novice workers, but not by the recruiters. These were:

* a facet of self confidence: to take risks;
* a facet of working with others: to 'read' what is happening in a group;
* a facet of communication: to communicate with someone older or more senior; and
* another facet of communication: public communication/ to present a professional image.

The meaning of these, along with the other facet descriptors, will be discussed.

Comparison of socially constructed personal skills with similar conceptions

Conceptions related to personal skills have been seen as an important part of education since the time of Aristotle, but, at least since the late nineteenth century, while such skills have been mentioned as important, they have not been an integral part of the assessable curriculum. Some early attempts to deal with this area will be discussed (such as Spearman 1927, 1950) particularly the affective domain taxonomy developed in the early 1960s by Krathwohl et al., (1964). It will be argued that a significant reason for the difficulties encountered by this taxonomy was the insistence upon separating cognition and affect.

A more useful way of structuring emotional intelligence is as a subset of socio-cultural understanding (which in turn is a subset of general ability). The approach of Mayer et al. (1997, 2000) can be seen as fitting this model where the conception of 'emotional intelligence' is described as a mental ability model. The strength of this model for the purpose of high stakes measurement will be shown by contrasting the construct with published mixed abilities models. The construct is useful because it is concerned with reasoning about emotions, not defining what those emotions or feelings should be or suggesting what kinds of attitudes are desirable. In terms of encapsulating the 'emotional intelligence' needed by novice workers, the Mayer et al. construct (1997) is useful as a model for the intrapersonal skills, but, having been designed for a different purpose, it does not adequately cover the interpersonal skills needed by novice workers.

It will be suggested that the Mayer et al. construct of emotional intelligence explores the distinction between cognition and personality in the realm of feelings. Another approach is to take the definition of 'intelligence' and acknowledge that it incorporates 'more' and 'less' cognitive components (McCurry & Bryce 1997). Gardner's (1993 a and b) work on multiple intelligences is useful here although it will be noted that these intelligences are not socially constructed (as discussed above), but viewed as 'biopsychological potentials' that can be destroyed or spared in isolation by brain damage. Although they are differently constructed, the conceptions of Mayer et al. (1997) and Gardner (1993 a and b) are very useful for looking at the alignment of cognition and feelings. Similarities and differences in the two constructs will be discussed.

This part of the discussion will conclude with a comparison of the Mayer et al. emotional intelligence, Gardner's personal intelligence and the socially constructed personal skills outlined above. I shall consider what needs to be done to shape the broader socially constructed personal skills into an ability model.

Issues concerning assessment

The paper will conclude with suggestions as to how these kinds of conceptions can be assessed in a rigorous manner. This will include discussion about the importance of definition so that there is a clear, shared meaning of what is meant, for example, by 'working with others'. The development of the construct is not yet complete, but the author will provide some examples of test items that model reasoning about intrapersonal and interpersonal issues, some of these use extracts from literature as stimulus material. The process of developing such items will be discussed, including the determination of correct answers by a process of reasoning.

References

A.C. Nielsen Research Services. (2000). *Employer satisfaction with graduate skills.* Canberra: DETYA.

Adelaide Declaration. (1999). *National Goals of Schooling for the 21st Century.* Internet address: www.curriculum.edu.au/mceetya/nationalgoals/natgoals.htm

Bar-On, R. (1997). *The Emotional Quotient Inventory: Technical Manual,* Toronto: Multi-Health Systems.

Gardner, H. (1993 a). *Frames of Mind,* 2nd edition, London: Fontana Press.

Gardner, H. (1993 b). *Multiple Intelligences,* New York: Basic Books.

Krathwohl, D., Bloom, B & Masia, B. (1964). *Taxonomy of Educational Objectives, Handbook II: The Affective Domain.* London: Longman Group Limited.

McCurry, D. & Bryce, J., (1997). *The School Based Key Competencies Levels Assessment Project.* Final Report, April 1997. Camberwell: Australian Council for Educational Research.

Mayer, J.D., Caruso, D.R. & Salovey, P. (1999). Emotional Intelligence Meets Traditional Standards for an Intelligence. *Intelligence 27* (4), 267–298.

Mayer, J. D. & Salovey, P. (1997). What is emotional intelligence?, in P. Salovey. & D Sluyter (eds), *Emotional Development and Emotional Intelligence.* New York: Basic Books.

Mayer, J.D., Salovey, P. & Caruso, D.R. (2000). Models of Emotional Intelligence, in R. Sternberg (ed.) *Handbook of Intelligence.* Cambridge: Cambridge University Press.

Spearman, C. (1927). *The Abilities of Man,* London: Macmillan & Co. Limited.

Spearman, C. & Wynn Jones, L. (1950). *Human Ability,* London: Macmillan & Co. Limited.

Wellington, J. (1994). How far should the post-16 curriculum be determined by the needs of employers? *The Curriculum Journal 5,* 3.

A Leader's Profile – *What attributes make an effective leader?*

David Rosete

Australian Taxation Office

David Rosete has some 10 years experience in both the private and public sectors with particular strengths in the filed of organisational psychology, human resource management and scientific research. Currently David works for the Australian Tax office as the Director of Organisational Development, which is an area specialising in implementing organisational development initiatives across the ATO, including assessment work, leadership development and change management. David has also worked as a casual lecturer and tutor at the University of Wollongong teaching in both the Department of Management and the Department of Psychology. Within the Dept. of Psychology David is currently carrying out his PhD research in the area of leadership effectiveness, emotional intelligence, personality and cognitive intelligence.

Abstract

There is growing evidence to suggest that today's executives are derailing from their career paths at record-high rates of up to 50 percent. This suggests that organisations need to focus more efforts on accurately identifying and selecting people with leadership potential (Corporate Leadership Council 2003; Blohowiak 2003). Furthermore, organisations rely too heavily on past performance as an indicator of future performance – this alone has been shown to be insufficient (Kesler 2002). This study presents evidence on instruments that can be used to identify leadership potential and supports the significant value of a holistic approach to assessment, involving a battery of tests (i.e. cognitive intelligence, personality and emotional intelligence) and multiple information sources, rather than any one single assessment instrument.

Introduction

While there appears to be an overabundance of leadership research we are still no closer to understanding what is required to be an effective leader (Kets de Vries 1994; Higgs and Rowland 2001; Hogan and Hogan 2001; Higgs 2003). Questions still remain around why clearly intelligent and experienced leaders are not always successful in dealing with environmental demands and life in general. Perhaps the answer rests with the new Neo-charismatic theories of leadership. Neo-charismatic theories of leadership explore factors that allow leaders to be extremely empathic with their staff, to have flourishing intimate relationships, and generally be thought of as charismatic by work colleagues. What is that makes one charismatic to others? Is it an innate trait that human beings have? Is it their general intelligence? Or is it another concept yet to be adequately defined, such as emotional intelligence (EI)?

Popular definitions of emotional intelligence refer to various things, including motivation, empathy, sociability, warmth, and optimism (Mayer 2001). EI approaches can be classified into two broad categories: (a) ability models (i.e., Mayer and Salovey 1997), and (b) mixed models. Ability models conceptualise EI in a similar way to cognitive intelligence (ie. IQ). EI is assumed to develop over time, be correlated with measures of IQ, and be measurable with a test based on performance (Ciarrochi, Chan & Caputi 2000). In contrast, mixed models of EI incorporate both non-cognitive models (e.g., BarOn 1997) and competency-based models (e.g., Goleman 1995). These mixed models tend to overlap or 'mix' with traditional models of personality and tend to utilise self-reports as their primary mode of assessment.

Regardless of which EI model one pursues, there seems to be a general acceptance that our relationships with people are highly dependent on our emotional, personal, and social dimensions of intelligence (Bar-On 1997). While there are still debates being carried out over what is emotional intelligence and how it is to be measured, the literature is somewhat clearer in highlighting the need for leaders to have more than simply cognitive intelligence or specific personality traits to succeed. Leaders are more likely to succeed in their relationships with others if they:

- understand their own feelings;
- are reflective of their own moods;
- are able to express feelings and empathy;
- can work cooperatively;

- can balance 'hard' and 'soft' decisions; and

- are able to promote social harmony.

Collectively, these behaviours or characteristics capture the essence of what the literature is calling Emotional Intelligence.

This paper examines how cognitive intelligence, personality, and EI relate to two measures of leadership effectiveness (the Performance Management System and 360 measures of leadership behaviours espoused within the workplace). While previous research (Rosete & Ciarrochi, in press) examined this relationship from an ability-based measure of EI, this paper focuses on the attributes that appear to influence effective leadership. These attributes include personality behaviours such as dominance and genuineness, an above average intelligence level, and emotional intelligence. Secondly, this paper explores the use of both self-report measures of EI and ability measures of EI, and what these EI measure can tell us about leadership effectiveness over and above the concepts of both personality and cognitive intelligence.

Participants

The sample consisted of 41 executives from a large Australian Public Service organisation with 24 (57.14%) respondents being male, and 18 (42.86%) female. Participants' ages ranged from 27 to 57 years old, and the average age was 42.24 years (standard deviation = 8.31). While 75 percent of participants had been with the organisation for 10 years or more (M = 15.56, standard deviation = 8.20).

Ability-based measure of emotional intelligence

The Mayer-Salovey-Caruso Emotional Intelligence Test, Version 2.0 (MSCEIT

V2.0; Mayer, Salovey, & Caruso, 2002) was used in this study. The MSCEIT V2.0 (Mayer et al., 2002) contains 141 items which are broken down into eight tasks, which are further divided into four branches of abilities including (a) perceiving emotion, (b) using emotion to facilitate thought, (c) understanding emotion, and (d) managing emotions. Mayer et al. (2002) reported reliabilities of $\alpha = 0.91$ for the full scale, $\alpha = 0.81$ for emotional management, $\alpha = 0.77$ for emotional understanding, $\alpha = 0.76$ for emotional facilitation, and $\alpha = 0.90$ for emotional perception. For this group, the mean total emotional intelligence score was 100.1 (standard deviation = 15.17).

Self-report measure of emotional intelligence

The Swinburne University Emotional Intelligence Test (SUEIT, Palmer & Stough 2001) was also used in this study. The SUEIT is a self-report measure consisting of 65 items. Items of the SUEIT ask participants to indicate the extent to which various statements (items) are true of the way they typically think, feel and act at work (on a five-point scale: 1 = never, 2 = seldom, 3 = sometimes, 4 = usually, 5 = always). The SUEIT report provides an overall EI score that indicates a participant's general workplace EI, and five sub-scale scores including (1) emotional recognition and expression (in oneself); (2) emotions direct cognition; (3) understanding of emotions external; (4) emotional management; and (5) emotional control. Palmer and Stough (2001) reported reliabilities of $\alpha = 0.91$ for the total EI scale, $\alpha = 0.77$ for emotional recognition and expression (in oneself); $\alpha = 0.70$ for emotions direct cognition; $\alpha = 0.89$ for understanding of emotions external; $\alpha = 0.83$ for emotional management; and $\alpha = 0.77$

for emotional control for executive groups. For this group, the mean total emotional intelligence score was 49.4 (standard deviation = 18.47).

Measure of personality

Participants completed the well-validated Sixteen Personality Factor questionnaire (16PF, Conn & Rieke 1998). The total scale contains 185 items and each subscale contains 10 to 15 items which form 16 primary factors. The primary factors are further analysed into five global factor scales that align with the five factor model of personality. These include extraversion (M = 4.3, standard deviation = 2.21), anxiety (M = 5.5, standard deviation = 2.01), tough-mindedness (M = 5.2, standard deviation = 1.98), iIndependence (M = 4.9, standard deviation = 2.04) and self-control (M = 4.44, standard deviation = 1.72).

Measure of cognitive ability

The Wechsler Abbreviated Scale of Intelligence (WASI – Psychological Corporation 1999) was used to measure cognitive ability. The WASI consists of four subtests: Vocabulary, Block Design, Similarities, and Matrix Reasoning. These subtests measure an individual's expressive vocabulary, verbal knowledge, visual-motor coordination, abstract conceptualization, verbal reasoning ability and nonverbal fluid reasoning. Analysis of the data provides three scores: a full scale IQ, a verbal IQ score, and a performance IQ score. The full scale IQ for this group was 118.3 (standard deviation = 10.43) mean verbal IQ was 117.2 (standard deviation = 12.31), and performance IQ was 114.9 (standard deviation = 10.67).

Measures of leadership effectiveness

Two measures of leader effectiveness (often used within the Australian Public Service) were used in this study:

- A performance management system; and
- 360 measures of leadership behaviours espoused within the workplace (also known as multi-rater measures).

The performance management system

The purpose of the Performance Management System is to evaluate an executive employee's performance in achieving agreed business outputs in the previous financial year (known as the '*what*' i.e. 'What has been achieved?') and to evaluate how the employee demonstrated the expected leadership behaviours in achieving those outputs (known as the '*how*', i.e. 'How has it been achieved? Did the employee model the core values of the organisation?'). The '*what*' and '*how*' evaluations highlight two separate but related aspects of an individual performance (Management Advisory Committee 2001). Both the 'what' and 'how' are rated on a five-point scale (1 to 5) by the participant's direct manager. Individuals are not rated on their innate abilities, knowledge or skills, but rather on how well these abilities, knowledge or skills have been applied in the achievement of business outputs over the financial year.

Multi-rater assessment

All participants were also asked to complete an on-line multi-rater (360°) instrument, called the *Perspectives on Executive Leadership Capabilities* (PELC). It comprises 40 behavioural statements that relate to the Australian Public Service Commission (APSC)

Leadership Capability Framework. Those core leadership capabilities include:

- *shapes strategic thinking* (i.e. inspires a sense of purpose and direction; focuses strategically; harnesses information and opportunities; and shows judgment, intelligence and commonsense);
- *achieves results* (i.e. builds organisational capability and responsiveness; marshals professional expertise; ensures closure and delivers on intended results; and steers and implements changes and deals with uncertainty);
- *cultivates productive working relationships* (i.e. nurtures internal and external relationships; values individual differences and diversity; guides, mentors and develops people; and facilitates co-operation and partnerships);
- *communicates with influence* (i.e. communicates clearly; listens, understands and adapts to an audience; and negotiates persuasively); and
- *exemplifies personal drive and integrity* (i.e. engages with risk and shows personal courage; commits to action; displays resilience; and can demonstrates self-awareness and a commitment to personal development).

Individuals completing the PELC were asked to rate their own leadership effectiveness. Direct staff and the individual direct manager also rated the person on the same criteria – that is, a manager and at least three subordinates (M = 3.6 subordinates per participant). No data on age or gender were collected to maintain the anonymity of direct managers and their participants' subordinates. While we would expect a positive correlation between individuals' performance ratings and their results obtained on the multi-rater questionnaire, the multi-rater offers us

further insight into leadership behaviours espoused by individuals, as it includes views from staff and manager. The PELC's scale reliability coefficient was 0.93.

Procedure

Participants were administered a battery of psychological tests (i.e., 16PF, MSCEIT, SUEIT and WASI). All participants were provided with information regarding the instruments, consent forms, a copy of the 16PF, and a copy of the SUEIT. Participants were also provided with either a paper and pencil version of the MSCEIT or computer access codes for completing the MSCEIT on-line. Participants were also administered the WASI. In exchange for their participation, individuals were provided with a confidential feedback report on their results for each of the instruments.

Preliminary analyses were conducted on the relationship between cognitive intelligence, personality, MSCEIT, SUEIT and the two measures of leader effectiveness (i.e., the performance management system and multi-rater measures of leadership). Rosete and Ciarrochi (2004) report much of these findings which focused on the MSCEIT, as it was found to be the best predictor of performance measures of leadership effectiveness. This paper focuses more on the unique relationships that cognitive intelligence, personality and self-report measures of EI have with the multi-rater measures of leadership behaviour.

Cognitive intelligence, personality, EI and leadership effectiveness

Pearson correlation coefficients were used to examine any relationships that may exist between IQ, the 16PF Five Global Factors, and the MSCEIT and SUEIT measures of emotional intelligence, with both measures of leadership effectiveness. These correlations are presented in Table I.

Table I Correlations coefficients between the MSCEIT, SUEIT, IQ, 16PF five global factors with leadership effectiveness measures

	SST	CPWR	EPDI	CWI	AR	PMR – What	PMR – How
Total MSCEIT score	.20	.34*	.37*	.16	.22	.02	.38*
Total SUEIT score	.59**	.51**	.50**	.45**	.34	.17	.27
Total IQ score	.12	.12	.29	.09	.39*	.04	.20
Extraversion	.49**	.47**	.29	.30	.09	.18	.15
Anxiety	-.63**	-.56**	-.53**	-.45**	-.39*	-.25	-.28
Tough mindedness	-.09	.04	.14	-.10	.23	.003	-.04
Independence	.42*	.32	.19	.30	.15	.29	.04
Self control	-.06	.14	.27	.01	.30	.19	.14

SST = shapes strategic thinking; CPWR = cultivates productive working relationships; EPD = exemplifies personal drive & integrity; CWI = communicates with Influence; AR = achieves results; (PMR) – What = performance management rating 'what' score, that is what business outcomes does the individual achieve; (PMR) – How = performance management rating 'how' score, that is how does the individual display leadership behaviours.
* Correlation is significant at the 0.01 level (2-tailed).
** Correlation is significant at the 0.05 level (2-tailed).

As can be seen in Table I, only the MSCEIT correlated significantly with the 'How' performance measure of leadership effectiveness (r (40) = 0.38, p < .05). These results support the notion that emotional intelligence is related to a leader's effectiveness in being able to achieve organisational goals. Cognitive intelligence, extraversion, anxiety and the SUEIT all appeared to impact core leadership behaviours as measured in the multi-rater instrument.

In respect to the SUEIT EI measure and the 16 PF, Table II below shows significant correlations between three of the global personality measures (i.e., extraversion, anxiety and independence) and the total EI score. Similarly, all five subscales of the SUEIT are related to extraversion.

The degree to which either personality or the SUEIT confounded each other in being able to predict any of the multi-rater factors was explored using stepwise regression. Independent stepwise regression analyses were calculated using each factor of the multi-rater leadership instrument (i.e., each multi-rater factor including *Shapes strategic thinking, Achieves results, Cultivates productive working relationships , Communicates with influence and Exemplifies personal drive and integrity*) as the dependent variables and any significant emotional intelligence, cognitive intelligence and personality factors as the predictors. We used an alpha criteria of .05 for variable entry and .10 for variable removal.

Anxiety emerged as the strongest predictor of the *Shapes strategic thinking* measure of leadership effectiveness (β = -0.653, p < 0.01, r = .43). This was followed by the SUEIT EI measure (β = 0.375, p < 0.05, r = .11).

The personality factor Anxiety also emerged as the strongest predictor of *Cultivates productive working relationships* (β = -0.569, p < 0.01, r = .32) and *Personal drive and integrity* (β = -0.547, p < 0.01, r = .30).

The SUEIT measure emerged as the strongest predictor of *Communicates with influence* measure of leadership effectiveness (β = 0.451, p < 0.01, r^2 = .20)

Table II Correlations coefficients between the SUEIT and 16PF five global factors

	SUEIT Total EI	SUEIT PERC	SUEIT PUOE	SUEIT PEDC	SUEIT PEM	SUEIT PEC
Extraversion	.68**	.53**	.41*	.34*	.55**	.47**
Anxiety	-.48**	-.18	-.41*	-.21	-.36*	-.45**
Tough Mindedness	-.256	-.36*	-.03	-.43**	.10	-.08
Independence	.35*	.36*	.25	.11	.25	.21
Self Control	-.07	-.219	.19	-.17	-.03	-.03

PERC = perceived emotional recognition and expression, PUOE = perceived understanding of emotions external, PEDC = perceived emotions direct cognition, PEM = perceived emotional management, PEC = perceived emotional control.
* Correlation is significant at the 0.01 level (2-tailed)
** Correlation is significant at the 0.05 level (2-tailed)

Finally, Anxiety emerged as the strongest predictor of *Achieves results* measure of leadership effectiveness (β = -0.389, p < 0.05, $\underline{r}$ = .15). This was followed by the IQ measure (β = 0.385, p < 0.05, $\underline{r}$ = .15).

Discussion

The results of the current study revealed that cognitive ability appeared to play a rather limited role in accounting for why some leaders are more effective than others. This isn't surprising as the average IQ of the executive group was at least one standard deviation above the mean. This supports previous research around leadership and intelligence which seems to indicate that leaders tend to have above average intelligence and it is other attributes that allow an executive to succeed (Bahn 1979). Thus once you become an executive, it is other attributes that allow you to succeed in comparison to fellow executives. Other research on the use of general cognitive intelligence tests argue that in fact it isn't as good as a predictor as many claim as the percentage of variance of an individual performance on the job that is explain is anywhere between 0 percent to 25 percent (see for example Goldstein, Zedeck & Goldstein 2002).

Our results did support the notion that leaders low on anxiety (i.e. resilient leaders) are more likely to behave calmly, consistently, and predictably. This we predict means that on-the-job performance is likely to be higher as these leaders are more likely to have stable interpersonal relationships and be efficient in crisis management (Corporate Leadership Council 2003). These findings support previous 16PF research which showed that those executives demonstrating high emotional stability (i.e. they were mature, faced reality or were calm) and low vigilance (i.e. they were trusting,

accepting of conditions, and easy to get on with) were people who generally exhibited leadership potential (Dee-Burnett, Johns, Russell & Mead 1997).

The findings of Rosete and Ciarrochi (2004) suggested that executives higher on emotional intelligence were more likely to achieve business outcomes and be considered as effective leaders by their subordinates and direct manager. Regression analysis revealed that emotional intelligence, specifically the capacity to perceive emotions, could predict effective leadership over and above personality and cognitive intelligence using an ability measure of EI (i.e. the MSCEIT).

As for the self-report measure of EI, there is little doubt that there is now substantial evidence to suggest that self-report measures of EI correlate with personality (Ciarrochi, Chan & Caputi 2000; Ciarrochi, Chan, Caputi, and Roberts 2001; MacCann, Roberts, Mathews, and Zeidner, in press). Such overlap may make interpretation difficult, and may suggest that a result involving self-reported EI may simply be replicating previous results using a similar personality measure. These results support previous research by David (2003) that self-reports, particularly the SUEIT, is substantially and significantly correlated with the big five factors of personality.

However, one can not immediately infer that self-report measures are not valuable. Indeed, as demonstrated in this study there appears some evidence of the SUEIT utility, and it can be distinguished from other personality constructs when predicting multi-rater leadership ratings. Specifically the SUEIT was able to predict at least one multi-rater measure of leadership effectiveness (i.e. communicates with influence) over and above personality and IQ.

In conclusion, the picture emerging from this study and the research of Rosete and Ciarrochi (2004) indicates that an effective leader is one who has above average intelligence, exhibits resilience (i.e. low anxiety), dominance, and emotional stability. In terms of EI, self-report measures of EI appear to correlate highly with personality and not with cognitive intelligence. Self-report measures also appear to offer some additional information over and above personality and IQ when predicting at least one multi-rater leadership behaviour. While performance based measures of EI appear to be unrelated to personality, correlate with IQ and be the best predict for performance measure of leadership effectiveness. In effect we see a role for cognitive intelligence, personality and performance based measures of EI tests in being able to identify leadership potential.

References

Bar-On, R. (1997). *The Emotional Intelligence Inventory (EQ-i): Technical Manual*. Toronto, Canada: Multi-Health Systems.

Blohowiak, D. (20030. *Filling the Leadership Pipeline*. Lead Well Institute.

Ciarrochi, J., Chan, A and Caputi, P. (2000). A critical evaluation of the emotional intelligence construct. *Personality and Individual Differences 28*: 539–561.

Ciarrochi, J., Chan, A., Caputi, P., & Roberts, R. (2001). Measuring emotional intelligence. In J. Ciarrochi, J. P. Forgas, & J. D. Mayer (eds), *Emotional Intelligence in Everyday Life: A scientific enquiry* (pp. 25–45). Philadelphia, PA: Psychology Press/Taylor & Francis.

Con, S.R., and Rieke, M.L. (1998). *The 16PF Fifth Edition Technical Manual*. Champaign, IL: Institute for Personality and Ability Testing, Inc.

Corporate Leadership Council. (2003). *Assessment Methods for Identifying Leadership Potential*. Catalog No CLC11JRJ55.

David, S. (2003). Relationship of a measure of perceived emotional intelligence to personality and impression management. Paper presented at the Australian Industrial and Organisational Psychology Conference, 28 June, Melbourne.

Dee-Burnett, R., Johns, E.F., Russell, M.T., and Mead, A.D. (1997). *16PF Human Resource Development Report Manual*. Champaign, IL: Institute for Personality and Ability Testing, Inc.

Goldstein, H.W., Zedeck, S., and Goldstein, I.L. (2002). *Is this your final answer? Human Performance 15*(1–2), pp. 123–142.

Goleman, D. (1995), *Emotional Intelligence: Why it can matter more than IQ*, Bloomsbury Publishing, London.

Goleman, D. (1998), *Working with Emotional Intelligence*, Bloomsbury Publishing, London.

Higgs, M. (2003). How can we make sense of leadership in the 21st century? *Leadership & Organization Development Journal 24*(5), 273–284.

Higgs, M. and D. Rowland (2000). Building change leadership capability: the quest for change competence. *Journal of Change Management 1(2)*, 116–131.

Kets de Vries. M. R. (1993). *Leaders, Fools, Imposters*. San Francisco, CA: Jossey-Bass.

Kesler, G. C. (2002). Why the leadership bench never gets deeper: Ten insights about executive talent development. *Human Resource Planning*.

MacCann, C., G. Mathews, Zeidner, M., and Roberts, R.D. (2003). *Psychological Assessment of Emotional Intelligence: A critical review of self-report and performance-based testing*. Princeton, NJ: Center for New Constructs,.

Management Advisory Committee (2001). *Performance Management in the Australian Public Service: A strategic framework*. Commonwealth of Australia: Australian Public Service Commission.

Mayer, J. D. (2001). A field guide to emotional intelligence. In J. Ciarrochi, J. P. Forgas, & J. D. Mayer (Eds), *Emotional intelligence in everyday life: A scientific enquiry* (pp. 3–24). Philadelphia, PA: Psychology Press/Taylor & Francis.

Mayer, J. D. and Salovey, P. (1997). What is emotional intelligence? In P. Salovey and D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3–31). New York: Basic.

Mayer, J. D., Salovey, P., and Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)*, Version 2.0. Toronto, Canada: Multi-Health Systems.

Palmer, B. and Stough, C. (2001), *Workplace SUEIT: Swinburne University Emotional Intelligence Test – Manual*, Organizational psychology Research Unit, Swinburne University, Australia.

Rosete, D., & Ciarrochi, J. (2004). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. Manuscript submitted for publication.

Wechsler Abbreviated Scale of Intelligence, (1999). *Manual*. The Psychological Corporation.

Using the MBTI® for Leadership Development in a Professional Services Environment

Sue White

Consultant

Sue White, B A, B Ed, M. Ed, is the director of *peoplechange pty ltd*. Originally with a background in education Sue has worked consulting to a variety of organisations, teams and individuals in the areas of change management, conflict resolution and leadership development for the last 12 years.

Introduction

This presentation will report on my experiences in a professional services environment as the program learning adviser to 125 managers as they embarked on in-house leadership development programs in the mid 1990s. It will also refer to other approaches that some parts of the organisation made to using Type in their training programs.

In the 1990s the 'learning organisation' was a fashionable and yet elusive construct in more than one organisation. The Karpin report (1995) revealed that within some Australian industries there had been very little structured management development. It seemed not to have been considered as cost effective or necessary in the Australian commercial environment (Industry Taskforce on Leadership and Management Skills 1995).

Indeed in this particular organisation there was some confusion between the provision of learning and the processes of reward and recognition. There was never any doubt as to the commitment to technical /professional learning inside the organisation. There are clear requirements set by external bodies in the management of risk to ensure technical competence. The need for provision of any other learning was considered by many to be dubious, regardless of the findings of the Karpin report.

Leadership programs

My role was to implement the new and structured leadership programs in one geographic location. This was seen as the organisation's commitment to being a learning organisation and to life-long learning. A necessary first step was to secure the commitment of senior management and to then foster the active participation of those selected for the programs. There was a significant financial investment in offering and running the programs. Venue hire, facilitator fees, travel costs and possible time lost when potential revenue could have been earned are not easily dismissed in the commercial world.

The MBTI® had already been selected as the tool for self-knowledge, prior to my appointment to the organisation. This decision became the motivating force in my desire to be accredited in an instrument I had scant knowledge of and had been introduced to in less than favorable circumstances some years earlier! Having had a *done to/done over/boxing* experience in a postgraduate university program I entered the accreditation process with some trepidation.

Self-knowledge was seen as a key element for change in the organisational culture. The spin-off for better client service became a factor that provided for better buy-in by some parts of the organisation and a commitment to the MBTI for use in people development. For me, self-knowledge, the key to self-mastery, one of Senge's (1990) mental models, must be alive and well to create the learning organisation. I brought this belief with me into the way I ran the leadership programs. If self-knowledge was the key then a full understanding of one's Type preferences was critical.

The formal leadership program consisted of a seven modules undertaken over a 15 to 18 month period. The people chosen to teach or facilitate each module were experts in their field. Many of them had international reputations and were published and acclaimed nationally at the forefront of management training.

Not all of them had the knowledge of group dynamics to work with the resistance that they would encounter from some of the managers.

For these leadership programs to be successful for the organisation and the individuals concerned, the natural skepticism of the training group needed to be harnessed to maximise their learning process. Comfort and relative ease in using what was for them such a suspect tool needed to be provided as well. These were the challenges!

Role of the program learning adviser

The role of the program learning adviser was new to the organisation and definitely to the managers themselves. Whilst the logistics function of the role could be understood it was not apparent to any of the groups at the beginning of their programs how another person could assist their learning process. The introductory sessions, headed by very senior management, spoke of the strategic need for new kinds of learning, especially action-learning. The groups were generally silent and acquiesced in the main activity of the session which was the completion of the Honey and Mumford learning style questionnaire followed by the issuing of the a large number of theoretical articles about action learning.

Later many mangers reflected on how the preference for introversion may not have been the one to exercise in this session. The fear of looking foolish also inhibited dialogue. It was felt that to ask questions was to reveal incompetence so it was best not to ask anything at all.

I understood my role to be one of making links, building relationships and making the connection between the learning and on the job experiences. I would also, make the link from one module to the next, informing the presenters of organisational issues that they might need to be aware of. I also remained mindful of where I saw the managers in their thinking processes and their readiness to try new experiences. In other words I was constantly evaluating the action learning process. I also made links between the managers and their sponsors if requested to.

The managers had been told that I could support them in their small group work if required for conflict resolution purposes. Their validated MBTI preferences generally meant avoiding conflict not resolving it! If the organisation preferred ISTJ preferences I adopted and extraverted NF preferences. As learning adviser I became a safe person to ask questions of and eventually assist in resolving conflict.

When briefing the presenter of the modules I asked them to share their Type with the group in order to reinforce the value of the instrument. It was of paramount importance to me that we made good use of Type, appreciating and valuing differences in style. Not all of our selected presenters knew their Type but, with the exception of one presenter, all either shared this information or allowed me to do the MBTI with them. One presenter so saw the value in reinforcing the learning in his module using Type that he became accredited in the instrument.

I have included the presenters' reported Types in the tables on the next page.

Program modules

During the first module, once some Type validation had occurred, we established the learning sets. According to Type difference and where the participants came from within the organisation, we created small groups of approximately five. It was intended that the group of managers at the more senior level take on an organisational issue that the managers believed needed a solution and, for the duration of the program, work proactively to find it. On the second module, for the junior managers group, this group elected to do a similar thing.

Each part of the organisation nominated a set number of managers to the programs and allocated sponsors. Sponsors took up their roles in a variety of ways. Some opened and closed modules and 'mentored' their people throughout the entire program. Sponsors were often invited to the closing sessions of module when the groups made presentations and attended the program celebration and review. Once a sponsor had been a member of the leadership program designed for the organisation's senior executive they were very keen to be involved. The MBTI was used on their program as well.

The program content was slightly different for junior and senior managers. The module outline for the more senior manager group (aged 28 plus and in the eighth year or more of their career) is outlined in the table on the next page.

Module	Content	Presenter's MBTI	Duration
1. Self and team	Team development Single loop and double loop learning Understanding self and others using the MBTI	ENFP or INTP	3 days
2. Strategic planning and change strategies	Planning models Scanning the environment Managing change for performance Organisation change strategies	XNTX	2 days
3. Managing the business's internal and external environment	Social political and economic issues facing the business world Challenges for Australian business Mega trends for the business of tomorrow The role of continuous improvement Managing risk and ethics	ESTJ	2 days
4. Leadership and culture	Leadership models Organisational cultures What successful leaders do Creativity and innovation Negotiation skills	ENFJ	4 days
5. Development of a sales culture	Introduction to marketing Strategic marketing concepts Developing marketing and selling skills Client service strategy	ENFJ	2 days
6. Presentation skills	Positive professional image Team presentation skills	ENTP	2 days
7. Final module	Closure, so what have we learned and what are the next steps?	INTP	1.5 days

For the slightly younger manager group (aged 26 plus and in the fifth or sixth year of their career) the modules were skills based, and operationally focussed. Each group of managers was divided into learning sets of four or five.

Module	Content	Presenter's MBTI	Duration
1. Self and team	Team development Single loop and double loop learning Understanding self and others using the MBTI	ENFP	3 days
2. Project management	Scoping a live project Using the organisation's CI processes	INFP	2.5 days
3. Communication skills	Verbal skills Listening skills Creating empathy Holding difficult conversations	ENFJ	2 days
4. Development of a sales culture	Introduction to marketing Developing marketing and selling skills Client service strategy	ENFJ	2 days
5. Presentation skills	Positive professional image Team presentation skills	ENTP	2 days
6. Negotiation skills	Understanding process Authenticity and influencing Building agreement and contracting	ENFP	2 days
7. Final module	Closure, so what have we learned and what are the next steps?	INTP	1.5 days

Group Types

Program 1 (senior)	ISTJ
Program 1 (junior)	ESTJ
Program 2 (junior)	ENTP
Program 2 (senior)	INTP
Program 3 (junior)	ESTJ

In all, over more than three years, nearly 125 managers were part of this program. We attempted to begin the lifecycle of a new group every six months. In the table above I have identified the group Types. There was no one modal Type. What was interesting was that in the younger groups E predominated, and as people remained in the organisation the I preference became the dominant one.

In true action learning we made modification to modules, presenters and venues as the groups moved through. The program for the younger managers was shortened from 15 months to 12 months, to ensure that all managers were still present locally by the end of the program and not seconded or promoted away.

Additional use of the MBTI

Two different parts of the organisation adopted the MBTI into their business unit education process. Stress was a topic that one business unit wanted to explore and people who had not been part of an official leadership program were then exposed to the MBTI in the exploration of stress at work. Many tales about group experiences rang

true for a significant number of managers. They then identified personal strategies for managing work stresses for themselves as managers and team leaders of others.

Another business unit introduced the use of the MBTI into their technical training in relation to building the client relationship. The rationale was that if the organisation's culture spoke with the language of one Type then it was important to learn to speak many Type languages to reach the widest variety of clients. In a two-day module using and experimenting with the functional pairs, the participants presented in preferred and non-preferred styles and gave each other feedback. This was considered fun and converted many who had been skeptical of the value the MBTI could add to their technical or client work.

Evaluation

It had been intended that as the program learning adviser I would formally evaluate the entire programs' effectiveness at their conclusion. However by then I had left the organisation to take on the challenge of change management using technology. Since leaving the organisation I have been asked to attend reunion dinners and breakfasts of learning sets and whole programs. It is now eight years since the first group went through their program and I know those members still employed in the organisation continue to meet once a year.

There were a number of events that occurred during the program that led me to believe that new leadership behaviours were emerging, old behaviours were changing and new knowledge was being translated into action. On an individual level there was no doubt that in terms of professional development most people believed that they were benefiting greatly.

Example one

During what should have been a problem-solving meeting, one manager opened up a leadership text she was carrying, read a portion aloud and then said, 'Have you noticed that we are not doing this?'

This action completely changed the dynamic of the meeting and then moved the relationships of the managers present to a more open one. They began to engage with each other much more readily. Another ESTJ takes action!

Example two

One group of managers arranged a meeting with a newly appointed senior executive to highlight issues they wanted to be part of solving. He came to the meeting, listened and stayed to eat lunch with the group afterwards.

They wanted to show that they could initiate action and work through possible contentious issues. This was considered unusual in the culture. The group who issued the invitation were INTP. The senior executive who stayed to engage with them was ENTP.

Example three

Two managers on a job late one night had decided to deliver a performance warning to one of their junior people. Instead, they got out their copy of *Type in Organisations*, reviewed their own Types and the section on contributions his Type could make to the team. They then reviewed the tasks he had been allocated. I was rung a couple of days later to be told there had been an amazing turnaround in his performance. 'This stuff does work!' The ISTJ and ESTJ had been struggling with the ENTP. The ESTJ drove the discussion with her colleague.

Example 4

One presenter questioned the interest of the group (almost to the point of offensiveness) because they were so quiet. In the morning tea break a group of them asked me to tell him they were introverts and were busy thinking and would talk later. 'Why doesn't he know this already?' said one INTP. An excellent question which demonstrated understanding of Type and the ability to use it!

References

Industry Taskforce on Leadership and Management Skills. (1995). *Enterprise Nation: Renewing Australia's managers to meet the challenges of the Asia–Pacific century.* (The Karpin Report.) Canberra: Australian Government Publishing Service.

Senge, Peter M. (1990). *The Fifth Discipline: The art and practice of the learning organization.* London: Random House.

A Self-Assessment Framework to Enhance an Individual's Development During the Executive Coaching Process

Patricia Pithers

Schroders Investment Management

Patricia Pithers holds a Bachelor of Business degree with a major in human resources. She is currently Human Resources Director for Schroder Investment Management Australia Ltd and has held that position for over eight years. Patricia has been in the human resources field for 18 years in a number of industries.

Executive coaching has become a key responsibility for Patricia, who now spends a great deal of her time coaching. This move toward executive coaching came when Patricia found that coaches external to the organisation were making little progress in helping executives change required behaviours for the long term. Her success in an internal coaching role comes from the interactions she has with coachees on a daily basis and the opportunity to observe them in the workplace.

Introduction

This session is intended to outline one approach to self-assessment, which has been used with many coachees, and to allow session attendees to experience the process first hand.

Providing individuals with a framework to assess themselves and to develop the ability to know and understand themselves is an important part of leadership development within an executive coaching program. All the reliability and validity of assessment instruments can be irrelevant if individuals do not see themselves in the results. The ability of individuals to understand themselves, to reflect, to be aware of the impact their behaviour has on others and to assert an honest commitment to change is the key to sustained behavioural transformation.

Executives and knowledge workers generally rise to the intellectual challenge. Using self-assessment in addition to the more common assessment tools such as MBTI or OPQ provides individuals with the opportunity to assess the impact their behaviour has on themselves and on others and the opportunity to decide whether they want to change those behaviours. The benefit for both the coach and coachee is the personal nature of the feedback, as it allows individuals to explore their values, attitudes and skills, in a non-judgmental environment where they alone make the rules.

There are many definitions of self–awareness, Daniel Goleman (1998) defines self–awareness as 'knowing one's internal states, preferences, resources and intuitions'. Nathaniel Branden (1997) describes it as 'a concern to understand the inner world of needs, motives, thoughts, mental states, emotions and bodily feelings'. Steven Covey (1989) has affirmed that 'self awareness enables us to stand apart and examine the way we see ourselves – our self paradigm, the most fundamental paradigm of effectiveness'.

All of these definitions are of course valid; we can probably relate to all of them in one context or another. In the executive coaching context, however, I have taken the Covey (1989) statement and built on it to help coachees to design their own self paradigm and determine the changes needed to improve their own leadership effectiveness.

The self-assessment framework, which is presented below, uses a series of questions and exercises which requires individuals to explore their purpose, values, attitudes, emotions, behaviour and skills as part of an Executive Coaching program. Throughout the process individuals are asked to list, discuss or evaluate their purpose, values, attitudes or skills and to describe the behaviours they demonstrate that support each of the items on their list.

Step 1

Determining the level of awareness
Establish the level of awareness and understanding the individual has of the values, thoughts, emotions and the accompanying behaviours, by working through the individual's goals, values, competencies and skills.

Step 2

Emotional awareness
Facilitate a process which allows the coachee to establish his or her level of emotional awareness and to articulate the behaviours which rise out those emotions.

Step 3

Cognitive awareness and awareness of the present and the external world
Facilitate a process, as required, which allows the coachee to determine the level awareness of the present, the external world and of the world of thoughts and memories.

Step 4

Focus on behaviour
Consolidate work done in steps 1–3 by exploring those behaviours required for success (in the appropriate context) by assessing existing behaviours and determining gaps. This step provides an opportunity for the coachee to see himself or herself as others do.

Step 5

Commit to change
Assist the coachee to determine which behaviours need to change and what he or she will need to undertake to achieve the required outcome.

This process can be time-consuming. It is an iterative process, which needs to be personalised to take into account the ability of the individual to develop the level of awareness required for sustained change. It is fair to say that some coachees are not ready or able to undergo such a process, or are not able to undertake the process in full. The process has proven highly successful for those coachees who are able to be open and honest with themselves, who are willing to try something a little different and not launch into specific skills development at the onset of a coaching program. Success also requires sensitivity and skill in the part of the coach. The coach must be able to say 'this is not working' or indeed refer the coachee to other professionals for issues outside of the scope of coaching, such as psychological or psychiatric states.

The process of coaching can be illustrated using this simple example.

M is a highly successful and capable salesman who has been promoted to a 'head of' position. In this new role M has five direct reports and has to 'manage and lead' this function for the foreseeable future.

M has excellent sales, strategy and people skills, but very limited day-to-day management skills, such as delegation, planning, time management or administration, and leadership skills, such as motivation, influencing, developing trust etc. He is very popular among the staff and his promotion is well received but there is work that needs to be done. After a short observation period it is discovered that M is highly strung with little emotional control, is very sensitive and shows almost no empathy.

M commences the coaching process. He is required to provide a list of values. One of his values is egalitarianism. He is then required to describe how he lives this value, specifically he is required to

describe the behaviours he exhibits which support this value. He is required to outline how living the value makes him feel. He has to describe how he draws fulfilment from living the value, and the impact that living the value has on others.

Through the coaching process M establishes that his value is demonstrated so strongly in many of his interactions with people at all levels including his subordinates. He also discovers that not all people have the same value and that they don't 'appreciate' what he is trying to do. He determines he needs to become more open to others and that he can get a lot more from others by being aware of who he is and learning about what makes others tick.

This process of self-knowledge, reflection and self-discovery can then be used to explore all aspects of the individuals leadership aptitudes, skills and competencies. Coachees are required to describe the type of leader they want to be, how they want to be perceived and how they need to behave in order to be perceived in this way. There is then a process of assessing the gaps. Many individuals find the process difficult and time-consuming. Those with the confidence and courage to persevere can start to quickly describe and develop those skills and behaviours that they need to be successful leaders. The leadership development process then follows.

Those individuals who are able to accurately describe their skills and behaviours, who are aware of their strengths and weaknesses, who are willing to learn from their experiences, who willingly seek feedback from managers, peers or subordinates and who are willing to reflect on these and take responsibility for their behaviour in an open constructive way find the process of behavioural change much more fulfilling, rewarding and successful.

References

Branden, Nathaniel. (1997). *The Art of Living Consciously: The power of awareness to transform everyday life*. New York: Fireside - Simon & Schuster Publishers.

Covey, Stephen R. (1989). *The Seven Habits of Highly Effective People*. New York: Fireside – Simon & Schuster Publishers.

Goleman, Daniel. (1998). *Working with Emotional Intelligence*. London: Bloomsbury Publishing.

Practical Applications of the Achenbach System of Empirically Based Assessment (ASEBA) for Ages 1.5 to 90+ Years

Thomas M. Achenbach

University of Vermont

Thomas M. Achenbach, Professor of Psychiatry and Psychology, is Director of the Center for Children, Youth, and Families at the University of Vermont Department of Psychiatry. A summa cum laude graduate of Yale, he received his Ph.D. from the University of Minnesota and was a Post-doctoral Fellow at the Yale Child Study Center. Before moving to the University of Vermont, Dr. Achenbach taught at Yale and was a Research Psychologist at the National Institute of Mental Health. He has been a DAAD Fellow at the University of Heidelberg, Germany; an SSRC Senior Faculty Fellow at Jean Piaget's Centre d'Epistémologie Génétique in Geneva; Chair of the American Psychological Association's Task Force on Classification of Children's Behavior; and a member of the American Psychiatric Association's Advisory Committee on DSM-III-R. He has given over 200 professional presentations and has authored over 200 publications, including *Developmental Psychopathology; Research in Developmental Psychology: Concepts, Strategies, Methods; Assessment and Taxonomy of Child and Adolescent Psychopathology; Empirically Based Taxonomy; Empirically Based Assessment of Child and Adolescent Psychopathology* (with Stephanie H. McConaughy); and Manuals for the *Child Behavior Checklist, Teacher's Report Form, Youth Self-Report*, and other standardized assessment instruments. Dr. Achenbach's honours include the Distinguished Contribution Award of the American Psychological Association's Section on Clinical Child Psychology and the University Scholar Award of the University of Vermont.

Introduction

This workshop presents instruments for assessing behavioural, emotional, and social problems and adaptive functioning. Developmentally appropriate instruments are designed for assessing preschool children, school-age children, adolescents, adults between the ages of 18 and 59, and adults over the age of 59. To obtain multiple perspectives, self-report forms are completed by the person who is being assessed and parallel forms are completed by people who know the person being assessed. The informants include parents, teachers, spouses, partners, caregivers, friends, and grown children of the person being assessed. ASEBA software systematically compares problems reported by each respondent. The systematic comparisons reveal similarities among and differences between problems reported by each respondent. The ASEBA also includes the *Semistructured Clinical Interview for Children and Adolescents* (SCICA) and the *Test Observation Form* (TOF), which documents problems observed during psychological testing.

We first present ASEBA assessment forms, scales for scoring problems and competencies, profiles for displaying item and scale scores, cross-informant comparisons, and multicultural applications. Thereafter, we illustrate applications of ASEBA instruments to particular cases.

ASEBA instruments for ages 1.5 to 5 years

For ages 1.5 to 5, the primary instruments are the *Child Behavior Checklist for Ages 1.5-5 and Language Development Survey* (CBCL/1.5-5-LDS) and the *Caregiver-Teacher Report Form*

(CTRF). The CBCL/1.5-5 includes 99 items that describe specific kinds of behavioural, emotional, and social problems that characterise preschool children. Parents and surrogates rate each item as *0 = not true, 1 = somewhat or sometimes true*, or *2 = very true or often true* of the child, based on the preceding two months. There are also open-ended items for describing additional problems, illnesses and disabilities, what concerns the respondent most about the child, and the best things about the child. The LDS requests information about risk factors related to language development, the child's use of phrases, and the child's vocabulary, based on a list of 310 words typical of children's early language.

The C-TRF is a parallel form completed by preschool teachers and daycare providers. It has many of the same problem items as the CBCL/1.5-5, but also has items specific to group settings in place of CBCL/1.5-5 items that are specific to home settings.

Scales for scoring the CBCL/1.5-5-LDS and C-TRF

The LDS is scored on normed scales for the number of vocabulary words and the length of phrases reported for the child. The problem items of the CBCL/1.5-5 and C-TRF are scored on the following empirically based syndromes: *Emotionally Reactive; Anxious/Depressed; Somatic Complaints; Withdrawn; Attention Problems*; and *Aggressive Behavior*. The CBCL/1.5-5 also has a *Sleep Problems* syndrome. The syndrome scales were constructed 'from the bottom up' by starting with data on problems reported for large samples of children and then using multivariate statistical analyses to identify sets of problems that co-occur.

The preschool forms are also scored on DSM-oriented scales constructed from the 'top-down' by having experts from many cultures identify items that they judged to be very consistent with diagnostic categories of the American Psychiatric Association's *Diagnostic and Statistical Manual* (the DSM). The DSM-oriented scales are: *Affective Problems; Anxiety Problems; Pervasive Developmental Problems; Attention Deficit/Hyperactivity Problems*; and *Oppositional Defiant Problems*. The empirically based syndromes and DSM-oriented scales are displayed on profiles in relation to percentiles and standard scores (T scores) based on national normative samples.

ASEBA instruments for ages 6 to 18 years

For ages 6 to 18, the *Child Behavior Checklist for Ages 6-18* (CBCL/6-18) is completed by parents and surrogates, while the *Teacher's Report Form* (TRF) is completed by teachers and other school personnel. The *Youth Self-Report* (YSR) is completed by youths to describe their own functioning. The problem items of the CBCL/6-18, TRF, and YSR are scored on the following empirically based syndromes: *Anxious/Depressed; Withdrawn/Depressed; Somatic Complaints; Social Problems; Thought Problems; Attention Problems; Rule-Breaking Behavior*; and *Aggressive Behavior*. The CBCL/6-18, TRF, and YSR are also scored on the following DSM-oriented scales: *Affective Problems, Anxiety Problems, Somatic Complaints; Attention Deficit/Hyperactivity Problems; Oppositional Defiant problems*; and *Conduct Problems*. The scales are displayed on profiles in relation to gender and age-specific percentiles and T scores based on national normative samples.

The CBCL/6-18 and YSR are also scored on competence scales for activities, social involvement, school, and total competence. The TRF is scored on

scales for academic performance and adaptive functioning. Like the problem scales, the competence, academic, and adaptive scales are displayed on profiles in relation to percentiles and T scores based on national normative samples.

ASEBA instruments for ages 18 to 59 and 60 to 90+ years

The *Adult Self-Report* (ASR) and *Adult Behavior Checklist* (ABCL) are used to obtain self-reports and informants' reports, respectively, of problems and adaptive functioning for ages 18 to 59. The *Older Adult Self-Report* (OASR) and *Older Adult Behavior Checklist* (OABCL) are similarly used to assess problems and adaptive functioning for ages over 59. The normative samples for the older adult forms included substantial numbers of people who were 60 to 90 years old, but the research samples included ages up to 102.

For ages 18 to 59, the syndrome scales are: *Anxious/Depressed; Withdrawn; Somatic Complaints; Thought Problems; Attention Problems; Aggressive Behavior; Rule-Breaking Behavior*; and *Intrusive*. For ages above 59, the syndrome scales are: *Anxious/Depressed; Worries; Somatic Complaints; Functional Impairment; Memory/Cognition Problems; Thought Problems*; and *Irritable/Disinhibited*. For ages 18 to 59, the DSM-oriented scales are: *Depressive Problems; Anxiety Problems; Somatic Problems; Avoidant Personality Problems; Attention Deficit/Hyperactivity Problems*; and *Antisocial Personality Problems*. For ages above 59, the DSM-oriented scales are: *Depressive Problems; Anxiety Problems; Somatic Problems; Dementia Problems; Psychotic Problems*; and *Antisocial Personality Problems*.

The adult forms also have items for assessing substance use and adaptive functioning, including *Friends;*

Spouse/Partner; Family; Job; Education; and *Mean Adaptive*. For ages above 59, there is also a scale for *Personal Strengths*.

Cross-informant comparisons

Hand- and computer-scored profiles for parallel forms can be visually compared to identify similarities and differences in scores obtained from self-reports and informants' reports. The ASEBA software prints side-by-side comparisons of item scores and histograms of scale scores obtained from all forms completed for an individual. To indicate how well particular respondents agree, the software prints *Q* correlations between item scores obtained from pairs of informants. It also prints *Q* correlations from large reference samples of respondents as a basis for determining whether agreement between particular informants is below average, average, or above average.

Cross-cultural applications

ASEBA instruments are the most widely used empirically based instruments in the world, with translations in 69 languages. Over 5,000 published studies by over 8,000 authors report use of ASEBA instruments in 62 cultures.

Case illustrations

The following cases illustrate how practitioners use ASEBA forms to obtain information from multiple informants in order to make diagnostic formulations, assess treatment needs, and guide interventions.

Kenny, age 30 months

Kenny was referred by his preschool teacher, who was concerned about his

lack of peer interaction and his talking to himself. CBCLs were completed by Kenny's parents and C-TRFs were completed by his two teachers. On the LDS, Kenny scored in the normal range for vocabulary and phrase length. However, comments written by his parents on the LDS indicated that he rarely used language to communicate with other children.

Scores from the four ASEBA forms were in the clinical range on the Withdrawn syndrome and the DSM-oriented *Pervasive Developmental Problems* scale. Scores were in the borderline range on the Emotionally Reactive syndrome. Ratings by both teachers, but not by Kenny's parents, were in the clinical range on the *Attention Problems* and *Somatic Complaints* syndromes. The side-by-side display of item scores made it easy to see which specific behavioural and emotional problems were endorsed by all informants and which were reported only by teachers or parents. ASEBA data and clinical observations strongly suggested a pervasive developmental disorder.

Alex, age 9.

Alex's parents sought an evaluation related to academic underachievement. His parents had recently separated and Alex spent time with each of them. Alex's main teacher reported the most problems, with scores in the borderline or clinical range on the *Anxious/Depressed, Withdrawn/Depressed*, and *Thought Problems* syndromes and on the DSM-oriented *Affective Problems* and *Anxiety Problems* scales. When Alex was interviewed with the SCICA, he obtained high scores on *Aggressive/Rule-Breaking Behavior* but low scores on anxiety and depression scales. Alex's *Kinetic Family Drawing* helped to explain the apparent discrepancy between his self-reports on the SCICA and reports by his parents and teachers. Alex's drawing depicted a

plane crash involving his family dressed in their Christmas clothes. After first depicting himself as a masked monster and then as a little boy trying to save the plane, he settled on depicting himself as a daredevil who had anticipated the plane's crash and bailed out in a dramatic plunge.

Jim, age 15

Jim was hospitalised for self-injurious behaviour, paranoid thinking, and school-refusal. Jim was previously diagnosed with attention deficit hyperactivity disorder, obsessive-compulsive disorder, and learning disability. He had been in an "emotional support" class for many years. ASEBA profiles were obtained from Jim's mother, three hospital staff members, and Jim himself. Ratings by all the adult informants placed Jim in the borderline or clinical range on the *Social Problems, Thought Problems, Attention Problems*, and *Aggressive Behavior* syndromes. Only the hospital staff rated Jim in the clinical range on the *Withdrawn* syndrome, whereas only his mother rated him as in the clinical range on the *Anxious/Depressed* syndrome. The most striking discrepancy was that Jim's YSR ratings yielded scores in the normal range on all syndromes except *Somatic Complaints* and *Social Problems*. Jim's case illustrates how ASEBA forms can document the presence of significant comorbidity, with elevations in syndromes that reflect both *Internalizing* and *Externalizing* problems in the same person. The relatively low YSR scores illustrate that youths may fail to report problems that are reported by others, although youths may also report problems that are *not* evident to others. In Jim's case, the adults' high ratings for *Thought Problems* suggest that Jim's failure to report such problems might be symptomatic of a severe mental illness.

Paul, age 4, and his parents

This case illustrates how practitioners can use ASEBA adult forms to facilitate work with parents of children presenting with behavioural and emotional problems. Paul Sternberg was referred for behavioural problems in nursery school. Lack of parental cooperation with behavioural management led the therapist to suggest that his parents each complete an ASR and an ABCL. Ms Sternberg's ASR yielded clinical range scores on the *Anxious/Depressed, Withdrawn*, and *Somatic Complaints* syndromes, on the DSM-oriented *Depressive Problems* scale, and on the *Spouse/Partner* and *Family* scales. The ABCL Mr Sternberg completed for his wife yielded clinical range scores on the *Attention Problems, Aggressive Behavior*, and *Anxious/Depressed* syndromes. Both adults concurred that Ms Sternberg was having difficulties with household responsibilities and was feeling depressed and overwhelmed. Mr Sternberg's ASR yielded clinical range scores on the *Aggressive Behaviour* and *Rule-Breaking Behavior* syndromes, the *Tobacco* and *Alcohol* scales, and the *Spouse/Partner* scale. Ms Sternberg's ABCL placed her husband in the clinical range on the *Aggressive Behavior* and *Withdrawn* syndromes and the DSM-oriented *Antisocial Personality Problems* scale. The therapist used this ASEBA information to recommend help for the Sternbergs in addressing their individual and marital problems in order to help their son.

References

Achenbach, T.M., Newhouse, P.A., & Rescorla, L.A. (2004). *Manual for the ASEBA Older Adult Forms & Profiles*. Vermont: University of Vermont Research Center for Children, Youth, and Families.

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Forms & Profiles.* Vermont: University of Vermont Research Center for Children, Youth, and Families.

Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Vermont: University of Vermont Research Center for Children, Youth, and Families.

Achenbach, T.M., & Rescorla, L.A. (2003). *Manual for the ASEBA Adult Forms & Profiles.* Vermont: University of Vermont Research Center for Children, Youth, and Families.

McConaughy, S.H., & Aschenbach, T. M. (2001). *Manual for the Semistructured Clinical Interview for Children and Adolescents.* (2nd ed.). Vermont: University of Vermont Research Center for Children, Youth, and Families.

McConaughy, S.H., & Aschenbach, T.M. (2004). *Manual for the Test Observation Form for Ages 2–18.* Vermont: University of Vermont Research Center for Children, Youth, and Families.

Assessment of Learning Disabilities in the Adolescent and Adult Populations: *Usefulness of the WAIS-III and the WJIII*

Noel Gregg

University of Georgia

Noel Gregg, Ph.D. is a Distinguished Research Professor at The University of Georgia, USA. She has published numerous books and articles in the area of learning disabilities and assessment. Her research has focused on adolescents and adults with learning disorders with a particular focus on reading and writing disorders, assessment, and accommodations. Currently she is involved in several international projects related to reading and writing disorders.

Abstract

Utilisation of broad and specific cognitive processing measures is essential to the assessment of learning disabilities. While technical manuals for psychometric measures provide reliability coefficients, this statistical information does not ensure that inferences drawn from test scores are defensible, specifically for the postsecondary population with learning disabilities. In addition, a comprehensive investigation of specific cognitive and linguistic processes (e.g., cluster and index scores) is critical for decision-making.

The purpose of this presentation will be threefold: (a) to discuss the construct validity of two commonly used broad cognitive ability measures with United States college students with (n=100) and without dyslexia (n=100); (b) to discuss the relationship between specific cognitive abilities as measured by the *Woodcock-Johnson III Cognitive Clusters* and the *Wechsler Adult Intelligence Scale III Index* scores; and (c) to discuss the cognitive and linguistic profiles of young adults with and without dyslexia.

The magnitude of difference between broad ability scores (e.g., full, composite, GIA) within and across two groups of young adults in the United States (with and without dyslexia) has serious implications for those individuals having access to accommodations.

Examination of the correlations between the *WAIS-III* Indices and the *WJ III COG* Clusters provides some interesting insight into the functional processing of each group (i.e., dyslexic/nondisabled) across tasks. Such information highlights the value of specific measures in the evaluation of the postsecondary population. Careful and informed selection of instruments by professionals is the safeguard that the rights of students are protected.

Program

Monday 19 July

9.00am	**Conference Opening**
	Presenter Professor Geoff Masters (CEO ACER)

9.30am **Plenary Session 1**   *Ballroom 3*

Presenter	Professor Thomas Achenbach (University of Vermont)
Title	**Psychopathology and Adaptive Functioning Across the Life Span:** *Top-down, bottom-up, multi-informant, and multi-cultural challenges and solutions*
Chair	Deirdre Morris (ACER)

10.30am **Morning Tea** (with sponsor exhibitions)

11.00am **Concurrent Sessions 1**

Session A	*Ballroom 1*
Presenter	Dr David Caruso (Harris McCully Associates Inc)
Title	**Applying the Mayer-Salovey-Caruso Approach to Emotional Intelligence**
Chair	Marian Power (ACER)

Session B	*Ballroom 3*
Presenter	Dr Ken Rowe (ACER) and
	Dr Kathy Rowe (Royal Children's Hospital Melbourne)
Title	**Developers, Users and Consumers Beware:** *Warnings about the design and use of psycho-behavioural rating inventories and analyses of data derived from them*
Chair	Kerry-Anne Hoad (ACER)

Session C	*Ballroom 2*
Presenter	Kevin Chandler (Chandler and McLeod) and Aimee Williamson (SHL)
Title	**Online Assessment:** *Current Issues and Future Views*
Chair	Ralph Saubern (ACER)

12.30pm **Lunch** (with sponsor exhibitions)
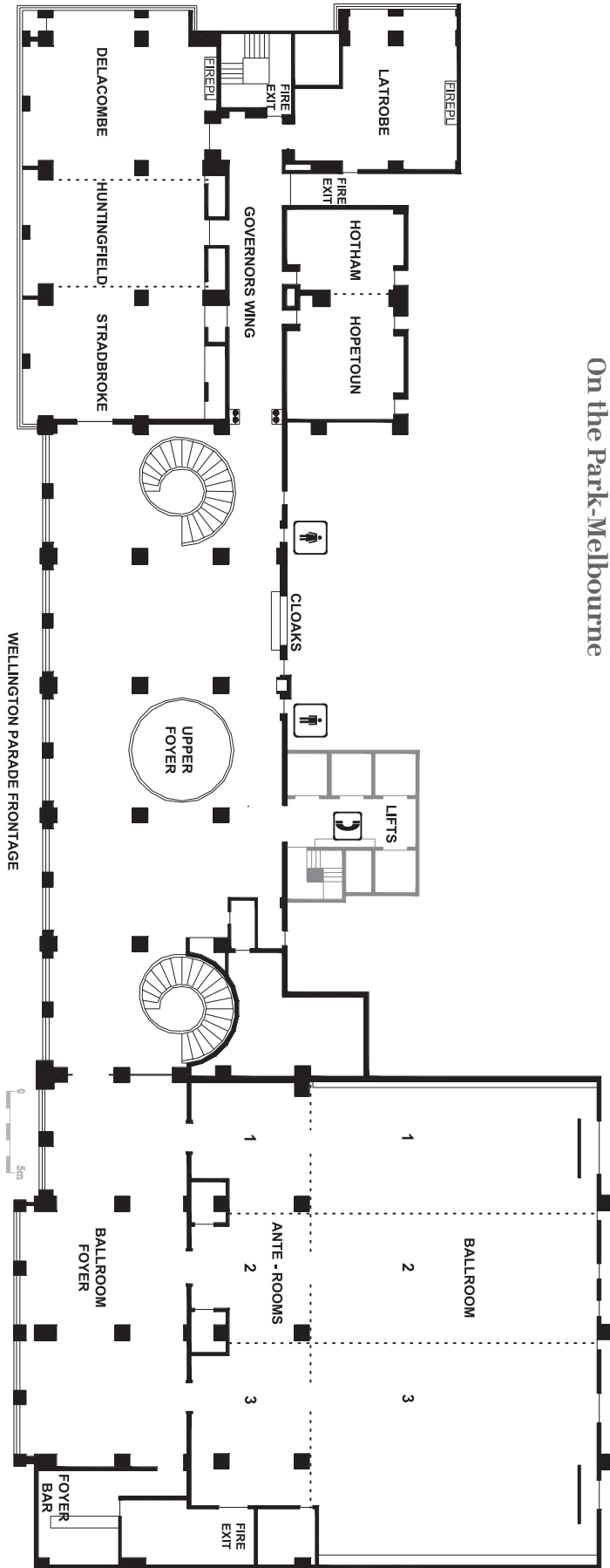
1.30pm   **Plenary Session 2**   *Ballroom 3*
        Presenter   Dr Katherine Hirsh (Sandra Hirsh Consulting (MBTI®))
        Title   **Viewing Organisations Through the Lens of Type:** *Applications of the MBTI® Tool*
        Chair   Marian Power (ACER)

2.30pm   **Afternoon Tea** (with sponsor exhibitions)

3.00pm   **Concurrent Sessions 2**

        Session A   *Ballroom 2*
        Presenter   Professor Gale Roid (Vanderbilt University)
        Title   **Uses and Advanced Interpretation of the Stanford-Binet Intelligence Scales, Fifth Edition (SB5)**
        Chair   Ralph Saubern (ACER)

        Session B   *Ballroom 3*
        Presenter   Dr Jennifer Bryce (ACER)
        Title   **Issues in Conceptualising and Assessing Emotional Intelligence**
        Presenter   David Rosete (Australian Taxation Office)
        Title   **A Leader's Profile –** *What attributes make an effective leader?*
        Chair   Kerry-Anne Hoad (ACER)

        Session C   *Ballroom 1*
        Presenter   Sue White (Consultant)
        Title   **Using the MBTI® for Leadership Development in a Professional Services Environment**
        Presenter   Patricia Pithers (Schroders Investment Management)
        Title   **A Self-assessment Framework to Enhance an Individual's Development During the Executive Coaching Process**
        Chair   Marian Power (ACER)

4.30pm   **Close of discussion**

7.00pm   **Conference Dinner at Hilton on the Park**

Assessing Intelligence, Emotion and Behaviour

Tuesday 20 July

9.00am   **Plenary Session 3**   *Ballroom 3*
Presenter   Dr David Caruso (Harris McCully Associates Inc)
Title   **An Intelligent Way to Assess Emotional Intelligence**
Chair   Kerry-Anne Hoad (ACER)

10.00am   **Morning Tea** (with sponsor exhibitions)

10.30am   **Concurrent Sessions 3**

Session A   *Ballroom 2*
Presenter   Professor Thomas Achenbach (University of Vermont)
and Professor Leslie Rescorla (Bryn Mawr College)
Title   **Practical Applications of the Achenbach System of Empirically Based Assessment (ASEBA) for Ages 1.5 to 90+ years**
Chair   Ralph Saubern (ACER)

Session B   *Ballroom 1*
Presenter   Professor Noel Gregg (University of Georgia)
Title   **Assessment of Learning Disabilities in the Adolescent and Adult Populations:** *Usefulness of the WAIS-III and the WJIII*
Chair   Marian Power (ACER)

Session C   *Ballroom 3*
Presenter   Dr Katherine Hirsh (Sandra Hirsh Consulting (MBTI®))
Title   **Creating the Teams You Want to Lead**
Chair   Kerry-Anne Hoad (ACER)

12.00pm   **Lunch** (with sponsor exhibitions)

1.00pm   **Plenary Session 4**   *Ballroom 3*
Presenter   Professor Gale Roid (Vanderbilt University)
Title   **Quality of Performance and Change-Sensitive Assessment of Cognitive Ability**
Chair   Marian Power (ACER)

2.00pm   **Afternoon Tea** (with sponsor exhibitions)

2.30pm   **Final Plenary Session**   *Ballroom 3*

Presenters   Dr Ken Rowe (ACER)
Kevin Chandler (Chandler and McLeod)
Wally Howe (Psychological Assessments Australia)
Title   **The Final Assessment – Lessons learned and bold predictions**
Chair   Deirdre Morris (ACER)

3.30pm   **Closing Address**   *Ballroom 3*
Presenter   Professor Geoff Masters (ACER)
Chair   Deirdre Morris (ACER)

Hilton
On the Park-Melbourne

DELACOMBE

LATROBE

FIREPL

FIRE EXIT

FIRE EXIT

HOTHAM

HUNTINGFIELD

GOVERNORS WING

HOPETOUN

STRADBROKE

CLOAKS

UPPER FOYER

WELLINGTON PARADE FRONTAGE

LIFTS

1

1

ANTE - ROOMS

2

2

BALLROOM

BALLROOM FOYER

3

3

FOYER BAR

FIRE EXIT

List of Delegates

Delegate	Organisation
Mrs Marlin Abd. Karim	Malaysian Examinations Syndicate
Professor Thomas Achenbach	University of Vermont, USA
Dr Sivakumar Alagumalai	Flinders University, SA
Professor Roslyn Arnold	University of Tasmania
Ms Asrijanty Asril	Educational Testing Center, Indonesia
Dr Kristine Bajuk	NSW DET
Mrs Hilary Barbour	TAFE NSW
Ms Julie Barclay	Melbourne Grammar School
Ms Margaret Bell	Nynan Public School, Nynan NSW
Ms Barbara Bini	Box Hill Tafe, VIC
Mr Geoff Bowser	CSU, NSW
Dr Ed Boyapati	RMIT
Mrs Louise Brown	Beth Rivkah Ladies College, East St Kilda VIC
Dr Jennifer Bryce	ACER
Ms Asmah bt Abd. Aziz	Malaysian Examinations Syndicate
Ms Adele Butler	ACER
Ms Kelly Callaghan	Sunshine Coast Grammar School
Ms Melissa Campbell	Catholic Education Office, Tweed Heads
Ms Helen Canny	Canny Alternatives Pty Ltd, VIC
Dr David Caruso	Harris McCully Associates Inc, USA
Dr Lynne Casey	
Mr Greg Cater	Somerset College, Mudgeeraba QLD
Mr Kevin Chandler	Chandler Macleod Group Ltd, NSW
Ms Tracy Clark	Notre Dame College, Shepparton VIC
Ms Mathilde Clements	SASI/Swinburne
Ms Nerida Collins	
Mr Mark Collis	Education Dept. QLD
Mrs Georgina Coombs	Firbank Grammar School, Brighton VIC
Mrs Jan Cowen	Ipswich Grammar School, QLD
Mr Piet Crosby	DECS, SA
Dr Nancy Cummane	
Ms Pamela Dawson	St Joseph's College, Pascoe Vale VIC
Mr Kris De Ross	One Test, QLD
Ms Jenny Dodd	Canberra Institute of Technology
Mr Peter Dodd	Radford College, Bruce ACT
Ms Lucia Doherty	Lucia Doherty Consulting, SA
Ms Linda Drake	DADHC, NSW
Ms Hilary Egan	
Mr Les Evans	Girton Grammar School, Bendigo VIC
Dr Helke Fahrner	Hampton Road Learning Centre, WA
Ms Janine Farrelly	JMF Psychology, VIC
Mrs Susan Feller	Joseph Varga School, Randwick NSW
Ms Geraldine Ferry Smith	University of Melbourne
Mrs Jennifer Finemore	Deakin University, VIC
Mrs Helen Gaudin	Melbourne Grammar School
Mr Ian Gelling	Scotch College, Hawthorn VIC
Mrs Bernadette German	Notre Dame College, Shepparton VIC
Ms Susanne Gilmour	Trinity Catholic College, Lismore QLD
Mr Kevin Graham	Canberra Institute of Technology

Delegate	Organisation
Professor Noel Gregg	The University of Georgia, USA
Dr Mark Gribble	Dept. of Defence, ACT
Ms Josephine Gullone	
Mr Maxwell Guppy	Cerebral Palsy League of QLD
Mrs Yvonne Hammer	St Paul's Grammar School, Penrith NSW
Ms Radmila Harding	Brighton Grammar School, VIC
Mrs Mawarni Hassan	Malaysian Examinations Syndicate
Ms Amanda Hawes	Cranbrook School, Bellvue Hill NSW
Mrs Donna Hayes	Catholic Education Office, QLD
Ms Michelle Henwood	Equalis, NSW
Dr Katherine Hirsh	University of Minnesota, USA
Mrs Kerry-Anne Hoad	ACER
Ms Fiona Hoath	St Paul's School, Bald Hills QLD
Ms Elaine Hosie	Melbourne Grammar School
Mr Wally Howe	Psychological Assessments Australia Pty Ltd
Mr David Huggins	Catholic Education Office, VIC
Dr Barbara Hummel-Rossi	New York University, USA
Ms Deirdre Jackson	ACER
Mrs Heather Jatan	Department of Education, TAS
Ms Jenny Johnston	AIST, TAS
Ms Kathleen E. Jones	Family Services, ACT
Ms Athanasia Koutsis	Brighton Grammar School, VIC
Ms Verna Kreibig	NSW DET
Ms Stephanie Lagos	Northern Migrant Resource Centre, VIC
Ms Marion Lambert	Korowa Anglican School, Glen Iris VIC
Assoc. Prof. Janice Langan-Fox	University of Melbourne
Mr Mark Lewis	Lewis & Lewis, VIC
Ms Sherry Livingstone	Education QLD
Mr Peter Luxton	QLD Studies Authority
Ms Jean MacKay	Catholic Education Office, Lismore
Mr Jeff Mason	ACT Education & Training
Professor Geoff Masters	ACER
Ms Verity Masters	IDSC, SA
Ms Natalie McColl	TAC, VIC
Dr Heather McCormack	
Mr Paul McGrath	Joseph Varga School, Randwick NSW
Ms Lee Minton	Redman & Associates, VIC
Ms Deirdre Morris	ACER
Mrs Lorraine Murphy	
Mr Damon Muttukumaru	Dept. of Defence, ACT
Ms Elizabeth Norris	Attainable Solutions, QLD
Ms Sharon O'Brien	Boys' Town Engadine, NSW
Ms Carole O'Leary	
Mr Lee Oliver	Lewis & Lewis, VIC
Mrs Mary Oski	Catholic Education Office, VIC
Dr Paul Paris	Salisbury High School, Salisbury North SA
Mrs Alexandra Parsons	Ivanhoe Grammar School, VIC
Ms Thelma Paull	
Dr Jiantoo Pei	Dept. of Defence, ACT

Delegate	Organisation
Ms Anne Peterson	ACER Press
Ms Patricia Pithers	Schroder Investment Management Australia Ltd
Ms Vivienne Powell	Methodist Ladies College, Kew VIC
Mrs Marian Power	ACER Press
Ms Sue Price	Bentleigh Secondary College, VIC
Mr Don Reid	NSW Dept. of Educ. & Training
Mrs Rhonda Robson	St Paul's Grammar School, Penrith NSW
Ms Keira Roffey-Mitchell	St Joseph's School, North Ipswich QLD
Professor Gale Roid	Peabody College of Vanderbilt University, USA
Ms Annmarie Rolls	ACER Press
Mr David Rosete	Australian Taxation Office, NSW
Ms Michelle Rowbotham	The University of Queensland
Dr Katherine Rowe	Royal Children's Hospital, VIC
Dr Ken Rowe	ACER
Ms Pam Russell	IARTV, VIC
Ms Liz Sadler	Health Insurance Commission, VIC
Mr Ralph Saubern	ACER Press
Mr Thomas Schick	Children's Court Clinic, NSW
Mr Rod Schultz	Education QLD
Mr Ricki Sharpe	Ricki Sharpe & Associates, NSW
Ms Venessa Sturre	The empower Group, VIC
Mrs Hendriastuti Suhono	Educational Testing Center, Indonesia
Ms Margaret Taylor	ACER
Dr Hanna Thomas	Workplace Research Associates, ACT
Dr Michael Thong	RMIT University
Ms Bronwyn Towart	Boys' Town Engadine, NSW
Ms Karen Underwood	Dept. of Education & Training, VIC
Ms Marina Veddovi	
Mrs Jenny Vincent Green	Firbank Grammar School, Brighton VIC
Mr Ken Wallis	City of Boroondara, VIC
Dr Tony Ward	Central Queensland University
Mrs Beth Watson	Firbank Grammar School, Brighton VIC
Dr Julie West	Workplace Research Associates, ACT
Ms Jennifer Wheeler	Dept. of Defence, ACT
Ms Philippa White	Workplace Research Associates, ACT
Ms Sue White	Peoplechange Pty Ltd, VIC
Mrs Joy Whitton	ACER Press
Ms Rebecca Wigg	ACER Press
Ms Aimee Williamson	Saville Holdsworth Ltd, NSW
Mrs Pamela Windsor	AIST, TAS
Ms Josephine Wintle	Network Recruitment Services, VIC
Mr Alec Young	ITE, TAS

(as at 3.00pm Monday 21st June)

International Test Users' Conference 2004