

# VALIDITY AS AN ACTION CONCEPT IN IO PSYCHOLOGY

CONRAD SCHMIDT

csch@rau.ac.za

Programme in Industrial Psychology  
Department of Human Resource Management  
University of Johannesburg

## ABSTRACT

In this article it is proposed that conventional conceptions of validity as applied in the field of Industrial and Organisational Psychological (IOP) assessment tend to emphasise technical aspects that result in an unintended separation of science and practice when implemented. An alternative conception of validity as an action concept is presented. It is noted that such a conception has been implicit in the field for some time and that the ideal of integrating science and practice, which stands so central to Industrial and Organisational (IO) Psychology, is promoted by it.

### Key words

Validity, action model, action science, statistical reasoning, professional judgment, region, psychological adequacy, actionability

It is widely acknowledged that validity represents the key issue in psychological assessment – and by implication IOP assessment (Muchinsky, 2003; Braun & Weiner, 1988). According to Shulz, Riggs and Kottke (1998) assessment and decision making in Personnel Psychology is based on the most rigorous procedures of positivistic science. It follows that the way in which the term validity has come to be understood in IO Psychology is, in the first instance, derived from positivistic conceptions of science.

According to this perspective the ideal form of knowledge is abstract, general and precise, as in the form of a mathematical equation to explain many different manifestations and cases (Kerlinger, 1978). Producing knowledge in this way ensures that it is objective – that is, not based on values, opinions and beliefs. Furthermore, by virtue of its nomothetic view of reality – as determined by general and abstract causal laws – positivistic procedures are aimed at producing knowledge that is universally valid and thus context independent (Neuman, 2003).

Schön (1983) pointed out that positivistic procedures tend to be based on a model of technical rationality – that is, one which applies science in a value-free manner. In the field of IO Psychology, Muchinsky (2004) has noted that conceptions of validity are typically conveyed in technical terms. One illustration of the technical emphasis in assessment is the advice that issues of test bias and test fairness should be kept separate – with the former viewed as a technical issue that resorts under the domain of science, and the latter as resorting under the domain of values outside the domain of science. A similar sentiment has been expressed regarding the issue of incorporating values in the concept of validity (Gregory, 2004) as argued by Messick (1995; 1980).

Historically the concepts of content, construct and criterion-related validity – referred to as the Trinitarian view of validity (Guion, 1980) – have captured the technical meaning of the term within IO Psychology (Cohen & Swerdlik, 2002; Shulz et al. 1998). According to Cohen and Swerdlik (2002) the acceptance of different kinds of validity is the prevailing view in Psychology, and has been so since the 1950s. They advise that the three approaches to validity assessment are not mutually exclusive, but should rather be thought of as types of evidence that – in conjunction with others – contribute to a judgment of the validity of a test. They offer the following advice in this respect:

All three types of validity evidence contribute to a unified picture of a test's validity, though a test user may not need to know about all three types of validity evidence. Depending on the use to which a test is being put, one or another of these three types of validity may not be as relevant as the next (p.155).

This appears to be the case in IO Psychology where criterion-related validity has been strongly emphasised (Shulz et al. 1998; Guion, 1991). Guion (1991) states:

Most research has concentrated on the evaluation of assessments – mainly tests, or predictors. If scores correlate with job behaviour of some sort – that is the criterion – then the assessment procedure is considered useful and valid, the level of validity being the correlation, or the validity coefficient (p.329).

The logic of criterion-related validity ... remains central to all personnel selection research (p.329).

It is apparent that the emphasis in IO Psychology on criterion-related logic and the expression of validity in mathematical terms is consistent with its framing as primarily a technical concept intended to convey accurate, precise, value-free and context-independent knowledge about the relationships between predictors and criteria, as expressed in a validity coefficient. The dominant action model of implementing such a conception of validity involve analysing the job, identifying predictors and criteria, testing many people with the same test for the same job, correlating the scores with a criterion and – if the correlation is satisfactory – selecting applicants with the best scores (Cascio, 1995; Guion, 1991). The references to selecting candidates with the best scores illustrate the inherent top-down logic of this model. It remains the current conceptual model used in IO Psychology, despite dating back several decades (Muchinsky 2003; Guion, 1976). With its emphasis on correlation and regression analyses, the model is in effect based on statistical reasoning.

Several considerations, however, may be identified that challenge the adequacy of the way in which validity has been conceptualised within IOP assessment, particularly as it relates to its implementation. These include, amongst others, the inherently normative nature of psychological assessment; conceptual challenges to the exclusive emphasis on criterion-related procedures to demonstrate validity; challenges to the assumption of the context-independence of predictor-criterion relationships, and the limited magnitude of predictive meta-analytic coefficients. Each of these considerations is discussed briefly below.

### *The inherently normative nature of psychological assessment*

A fundamental characteristic of psychological assessment, according to the Health Professions Council of South Africa (2001), is its intrinsic nature as a psychological act, the practice of which is restricted to those appropriately trained in the empirical behavioural scientific epistemology that represents the basis of professional education in this field.

In response to the technical rational model and from the perspective of applying science in practice, action theorists have drawn attention to the inherently normative dimension of action (Reason & Bradbury, 2001; Kemmis, 2001; Argyris, Putnam & Smith, 1985; Habermas, 1972). This refers to the premise that all deliberate action is aimed at producing certain outcomes rather than others. The implication of this perspective is that issues of value are inherently present when action is taken. By definition this would also apply to IOP assessment. Gregory (2004), for example, states that

...everyone can agree on one point: psychological measurement is not a neutral endeavour, it is an applied science that occurs in a social and political context (p.115).

Schön (1983) argued that issues of central concern in professional practice do not present themselves as neatly defined problems free from value considerations. When the principles guiding practice are predominantly of a technical nature, the risk exists that issues of value at the very core of professional practice are relegated to fall outside the scientific epistemology of the professional discipline itself. This, according to Schön (1983) gives rise to a typical dilemma faced by professional practitioners in general, namely that the definition of rigorous professional knowledge in which they have been schooled, excludes phenomena they experience as being central to their practice.

The implication is that the model of technical rationality that underpins the prevailing concept of validity does not account for the normative dimensions inherent to the psychological act of assessment. These dimensions are typically considered to be dealt with more effectively outside the realm of science through the specification of ethical codes (Gregory, 2004; Muchinsky, 2004; Anastasi & Urbina, 1997). However, the unintended consequence that tends to follow is that a gap develops between science and practice.

The implication is that if validity is central to IOP assessment and if theory and practice is to be integrated, a conception of validity is required that acknowledges the implicit values underlying action and the knowledge informing it (Argyris et al. 1985).

Conceptual challenges to the exclusive emphasis on criterion-related procedures to demonstrate validity in IOP assessment

In a review of changing conceptions of validity, Shulz et al. (1998) identified three stages that characterise the evolution of thought regarding its nature. In Stage 1 the emphasis was on defining the concept, resulting in definitions of different, independent types of validity. Emphasis was further placed on the importance of validating tests through external criteria. Accordingly a correlation coefficient between a test score and some external criterion was often presented as adequate evidence of a test's inherent validity.

Stage 2 thinking was characterised by viewing validity as a property of a test, with the implication that different types of validity could exist independently and that only certain types of validity needed to be shown when testing for different purposes. The existence of different types of validity was solidified in the "Trinitarian" doctrine (Guion, 1980). Not only did this doctrine enshrine validity as a property inherent to a test, but professionals were given the option of different methods of test validation:

Given a choice of validity types, it is natural that those most concretely defined and most simply obtained would be selected for practical applications. Consequently, procedures associated with content validation (asking experts if the items tap the construct of interest) and criterion validation (producing a correlation coefficient between the test and a selected criterion score) were much easier than attempting to

achieve construct validation that required launching a longitudinal attack involving multiple approaches. Either a coefficient of agreement among expert judges (content validity) or a correlation coefficient between test scores and a desired outcome (criterion-related validity) was presented and we slept soundly (Shulz et al. 1998, p.5).

In Stage 3 thinking, the conception evolved into a view that validity is not a property of a particular method of assessment, but rather of the appropriateness, meaningfulness and usefulness of the specific inferences made from test scores, reaffirming an earlier view articulated by Cronbach and Meehl (1955). They stated:

In one sense it is naïve to inquire "Is this test valid?" One does not validate a test, but only a principle for making inferences (p.297).

Guion (1991) re-affirmed this perspective by noting:

In any approach to validation, it is important to recognize that validation and validity refer to inferences drawn from data (scores), not to the predictors... It is not the predictor that is validated in empirical hypothesis testing... What is validated is the hypothesis that criterion performance can be inferred from the scores (p.350).

According to Huysamen (2002) the emphasis on validating the uses and interpretations of test or assessment results (rather than the test itself or the assessment procedures), was endorsed in the 1999 APA standards for educational and psychological testing. It is apparent that this perspective represents a significant shift away from the practice of viewing validity as an inherent property of the test (as the tendency has been in IOP assessment), towards an emphasis on the way test results are used.

Later views emphasised that validity is regarded as *something that is inferred, not measured*, and as something that is judged as adequate, marginal or unsatisfactory. Furthermore, validity has come to be regarded as a *unitary concept* based on evidence that includes the traditional categories of content, criterion and construct-related evidence. It became acknowledged that the evidence required to demonstrate validity may be accumulated in several ways, including all relevant data or facts as well as theoretical rationales or arguments that integrate such facts into an overall justification of test-score inferences.

Although the evidence may be obtained in many ways, it is the degree to which evidence supports the inferences made from scores that is considered to be the overarching essence of validity, with construct validity being regarded as the essential organising concept (Huysamen, 2002; Schulz et al. 1998). According to Schulz et al. the most direct implication of this conception of validity is that

...single points of evidence related to either content or criterion-related validity can no longer be offered as stand-alone indicators of measurement adequacy (p.5).

While it has been argued that construct validity represents the overarching organising concept in the unitary framing of validity, Guion (1991), from an IO Psychology perspective, voiced dissatisfaction with its adequacy as the appropriate organising concept. He draws distinctions between construct validity and job-related validity on the grounds that it is one thing to focus on describing a construct accurately, but quite another to make predictions on the basis of it. According to Guion construct validity is typically associated with psychometric research, where the concern is to confirm or disconfirm inferences regarding the meaning of scores, while job-relatedness is concerned with relational propositions about the attributes being measured and involves the testing of predictive hypotheses.

While the term “job-relatedness” may at first glance appear to be equivalent to the term “criterion-related/predictive validity”, Guion argues that job-relatedness need not be expressed in terms of a quantitative coefficient – such that criterion-related validity of the predictive type need not be a requirement, although it may include a consideration of it. Guion argues that informed professional judgment that follows the logic of criterion-related validity is the basis of job-related validity. However, by arguing for this distinction independent types of validity are implied – thus detracting from its nature as a unitary concept and opening the door once again to separate theory from action within IO Psychology, as is evident in his statement:

From a practical point of view, evidence of a predictor’s job-relatedness is more important than its psychometric validity (Guion, 1991, p.379).

An even broader view of validity has been advocated by Messick (1995, 1988, 1980) and Cronbach (1988). According to this perspective an appeal to empirical validity is not sufficient in justifying decisions based on test scores. Validity is also seen to encompass an evaluation of the value implications of both test interpretation and test use, in addition to an appraisal of its technical psychometric properties. Accordingly the interpretability, relevance and usefulness of scores, their value implications and functional worth as a basis for decisions, as well as their social consequences become key validity issues. In this broadened view, validity is conceived of as having both an evidential and a consequential basis (Messick, 1988).

The most significant implication of this view is that the scope of validity is broadened beyond a consideration of technical factors to encompass the normative dimension of action, and to include social consequences related to issues of justice. Although not all theorists agree with this broader interpretation, Cronbach (1988) points out that

...you... may prefer to exclude reflection on consequences from the meanings of the word “validation”, but you cannot deny the obligation (p.6).

Despite the above conceptual developments, it appears that they have not been incorporated into IOP assessment at the action level. For example, while the view of validity as a unitary concept has been advocated, an emphasis on the conceptual distinction between different types of validity evidence is still prevalent in texts and guidelines on assessment (Gregory, 2004; Huysamen, 2002; Cohen & Swerdlik, 2002; SIOPSA, 1998; Anastasi & Urbina, 1997; Murphy & Davidshofer, 1994).

In this regard Shulz et al. (1998) note:

While calls for the abandonment of the Trinitarian view of validity have been sounded for more than two decades (e.g. Guion, 1980; Landy, 1986; Schmidt & Landy, 1993; Tenoppy, 1977), in practice this view of validity still predominates. A logical question is *why?* (1998, p.1; italics added).

Critics of the traditional approach to validity point out that the conception of validity as manifesting in different types amounts to oversimplification, which can have a number of problematic consequences – one being that test users may focus on a single or small set of “validities”, rather than on the specific inferences they intend to make from the scores. Another consequence is that once evidence of a certain type of validity (for example, criterion-related validity) has been obtained, it may be regarded as being sufficient – when this evidence is forthcoming, the practitioner is considered as having been relieved from the responsibility of further inquiry. The criticism is that it amounts to selective reliance on one or a limited kind of validity evidence that, in turn, amounts to an over-generalisation (treating one kind of validity as representing validity in its totality) (Messick, 1988). Having

regard to the emphasis placed on validity as a unitary concept, it is clear that practice based on such a point of departure has become discredited from a conceptual point of view.

Little advice, however, appears to be forthcoming on how to act consistently with a concept of validity that is unitary in nature and encompasses more than technical aspects. As pointed out by Shulz et al. (1998) and noted much earlier by Schön (1983) the de facto situation appears to be that despite much criticism, the model practitioners are trained in, advocates a technical, value-neutral orientation and continues to dominate the field.

#### *Challenges to the assumption of the context-independence of predictor-criterion relationships*

Historical developments in criterion research have been characterised by efforts to define criteria in accordance with the knowledge requirements of empirical behavioural science, namely accuracy, precision, quantification and context-independence, so as to provide general statements about the relations between predictors and criteria (Austin & Villanova, 1992). The more recent emphasis on validity generalization in IO Psychology (Dunnette, 1998) illustrates the search for predictor-criterion relationships that hold across contexts. However, according to Austin and Villanova (1992) factor-analytic studies have borne out that criteria are dynamic, multidimensional, situation specific and complex in almost every case. These findings stand in sharp contrast to the assumptions of linear causality and precise, general and stable laws related to predictor-criterion relationships that operate across contexts – as encountered in traditional conceptions of validity.

One approach to dealing with this problem is to consider it the task of the practitioner to contextualise the information in the individual, situation-specific case (Cohen & Swerdlik, 2002; Anastasi & Urbina, 1997; Guion, 1991). Important, however, is the implication that action is then taken on the basis of practitioner judgment.

#### *The limited magnitude of predictive meta-analytic coefficients*

While the importance of accurate and precise data is espoused under the traditional conception of validity, results of meta-analytic studies indicate that at the core of the phenomenon of interest, greater unexplained than explained variance exists.

Oultz and Zedeck (in Campion, Oultz, Zedeck, Schmidt, Kehoe, Murphy, & Guion, 2001) point out that validity coefficients typically range between 0.20 and 0.50, accounting for between 4% and 25 % of the variance in the criterion. They state:

This means that 75% to 96% of the variance in the criterion is not accounted for. Yet, strict rank-order selection utilises predictor scores as if they account for the total variance in job performance (2001, p.152).

This statement illustrates two fundamental inconsistencies when acting purely according to a statistical reasoning model: Firstly, instead of acting on precise knowledge, action is taken on imprecise data that contains substantial unexplained variance. Secondly, the implementation of these results with their inherent limitations typically (because the implicit advice is to take the person with the highest score) proceeds as if they do not exist, on the basis of the logic that over time better decisions will be made:

We theorise that in the long run the mean criterion performance of samples selected via strict rank-order will be higher than those selected by any other method (Campion et al. 2001, p.152).

From an ethical perspective, however, the degree of error in the data becomes contentious because of the potential negative psychological impact it can have on individuals. The implication is that the actions of practitioners should account for and be consistent with limitations in the technology. In this

respect, the rigorous monitoring of error becomes central to the ethical stewardship of the practitioner. Failure to do so would in effect amount to treating the data as if there were no errors, thereby neglecting ethical stewardship. Importantly, it seems possible that the correct implementation of the technology on the basis of statistical reasoning could result in such neglect of ethical stewardship.

In the IOP assessment literature practitioner judgment is typically advised to compensate for limitations in the technology (Guion, 1991). In this way it is possible to identify strategies based on statistical reasoning and practitioner judgment respectively, as the dominant action models in IOP assessment. However, the reliance on professional judgment threatens the ideal of scientific objectivity (Highhouse, 2002; Dahlstrom, 1993) and in effect represents a circular argument: Practitioner judgment is to compensate for limitations in the technology while the limitations of practitioner judgment are to be remedied through objective procedures.

A well-respected author in the field of IO Psychology, Muchinsky (2004, p.207), notes that there has been much reference to the "scientist-practitioner gap" over the years and that this gap is particularly evident when it comes to implementing science. He states:

One major component of the gap is the issue of implementation. For the most part, scientists are relatively unconcerned with how their theories, principles, and methods are put into practice in arenas outside of academic study. For the most part, practitioners are deeply concerned with matters of implementation because what they do occurs in arenas not created primarily for scientific study (p.208).

Muchinsky points out that issues of implementation are rarely discussed in academic journals. He argues that it is of great importance that the basis of the scientist-practitioner gap be examined in order to better understand how it can be narrowed. Having regard to the above considerations, it appears that some evidence exists to suggest that the technical emphasis in conceptions of validity, as well as the tension that exists between statistical and professional models of action in IOP assessment may contribute to this gap. It is proposed that the reframing of validity as an action concept could contribute to addressing the scientist-practitioner gap referred to by Muchinsky.

#### **Reframing validity as an action concept**

It is apparent that the contours of a conceptualisation of validity in action terms have been evident in the field for some time. The assessment literature abounds with statements advocating that the effectiveness of tests are dependent on how they are used (Gregory 2004; Cohen & Swerdlik, 2002; Anastasi & Urbina, 1997; Guion, 1998, 1991; Matarazzo, 1990; Van den Berg, 1988). Muchinsky (2004), however, states:

A typical treatise on psychological testing is replete with concepts of a highly technical nature, such as statistical indices, estimates of reliability and validity, cut scores, sample sizes, measurement error, and so forth. What is conspicuously absent from such discussions are the human emotions associated with testing, from its creation, interpretation and consequences of its use. It has long been recognized that tests are but tools to help us make better decisions. What hasnot been acknowledged as extensively are the emotions associated with people drawn into the web of psychological assessment (p.206).

From the above it is evident that the framing of validity in technical terms only, ironically results in gaps in terms of accounting for the psychological aspects associated with its implementation.

Related to the emphasis on the use of tests by the practitioner is the way in which assessment information is used within the organisational context. Guion (1998) points out that deviations from the recommended scientific model of assessment based on criterion-related validity strategies (or, as it has been referred to in this article – a model based on statistical reasoning) is common in practice. He notes that such deviations may derive from limitations in terms of the user-friendliness of the model, given prevailing views that HR should operate as a business and consider other parts of the organisation as "customers" who can go to someone else for the service they need. He notes:

The system of using assessment results is perhaps more in need of evaluation than the assessment methods themselves... persistent harping on validity coefficients and utility analyses without clear, business-oriented evidence of profitability can lead to trouble (1998, p.359).

These comments suggest that IOP practitioners may unintentionally be contributing to pressures that undermine their professional actions (by virtue of the statistical terms in which the knowledge they produce, are shrouded). In addition, it may reasonably be expected that the legal implications of assessment may be experienced as so complicated and removed from the reality they deal with, that customers on the receiving end prefer to distance themselves from it. It appears that IO psychologists in industry are frequently put under pressure by line managers to reveal scores, without referring to the professional judgment of the psychologist. The obvious implication of such a practice is the risk that untrained persons then carry out psychological acts. It does not seem far-fetched to suggest that instances such as these illustrate the consequences of an epistemology where validity is viewed in technical terms only.

IO psychologists, perhaps more so than psychologists in other specialised areas, function in a larger context that impacts on the assessment process. From a broader perspective it can therefore be argued that the organisation is also ethically bound to assessment principles by virtue of its use of and reliance on such information. Acknowledgement of this principle would make it possible for the above risk to be addressed in a systematic and rigorous way, by virtue of all users (including organisational decision makers) being subject to the requirements of using assessment information validly. This would also be consistent with a view expressed by Guion (1998, 1991), namely that selection procedures should be evaluated according to how they are used within the context of a broader organisational focus.

If assessment procedures are to be evaluated in terms of the way in which they are used, it follows that a concept of validity is required that incorporates action considerations.

#### *Epistemological considerations*

From an epistemological point of view, the tension between statistical reasoning and professional judgment models of action in IOP assessment reflects a tension between methods associated with two levels of scientific knowledge, namely the empirical-analytical (emphasising statistical reasoning) and hermeneutic-interpretive (emphasising practitioner judgment). It is, however, apparent that each model has limitations in terms of the standards required by the other. For example, limitations of the statistical reasoning model include, amongst others, a lack of emphasis on the unique individual, limited predictive coefficients, the risk of atheoretical implementation of knowledge and limitations in providing context-specific knowledge. On the other hand, the limitations of professional judgment models include, amongst others, the risk of shoddy, untested practice and the absence of an adequate scientific model guiding practitioner judgment (Schmidt, 2006; Highhouse, 2002; Guion 1991).

The respective weaknesses of these action models could be said to represent a dialectical tension (Neuman, 2003), requiring adaptive integration or synthesis at a different level of inquiry (Argyris & Schön, 1996). The philosophy of science literature indicates that dialectical thought is associated with the critical level of knowledge (Neuman, 2003; Gordon, 2001; Snyman, 1993; Romm, 1993). At this level, ongoing critical reflective inquiry into the way in which knowledge is created and implemented at the action level becomes the key methodological vehicle (Argyris et al. 1985). The normative "action nature" of IOP assessment is seen to lend support to the proposition that an appropriate epistemological frame of reference for guiding practice will necessarily extend to the critical level of inquiry, allowing for the incorporation of normative considerations within a scientific framework.

A framework of this nature, that could prove helpful in respect of the goal of integrating science in practice in IO Psychology, is that of action science (Argyris et al. 1985; Argyris & Schön, 1974).

#### *Action science as frame of reference in IOP assessment*

Action science represents an epistemology characterised by an emphasis on high degrees of rigour in the practice context, through the enactment of core scientific principles in the actions of practitioners. Along with other approaches in the critical tradition, action science recognises that it is not possible to implement science in a value-neutral way, given the inherently normative nature of action (Denzin, 1994). It emphasises, for example, that the implementation of the control strategies inherent to the empirical behavioural science methodology is not a neutral act (Argyris, 1993; Kipnis, 1987). Following this line of reasoning it takes the position that action needs to be evaluated in terms of the values it claims to serve and that this is possible to achieve through an accountable scientific model based on critical-reflective inquiry (Kemmis, 2001; Habermas, 1972).

From a mainstream science point of view, rigour is reflected in high degrees of completeness, accuracy and precision to be achieved through the use of clearly defined procedures (Friedman, 2001; Argyris, 1980). From an action science perspective the notion that an increase in rigour is accomplished by an increase in precision, is tempered by the nature of the action context. This context is characterised by multiple interacting forces, many of which the practitioner has no control over. That this is true of the assessment context has been noted by several authors (Gregory, 2004; Kerlinger & Lee, 2002; Guion, 1998, 1991). For example, not all information about the individual is typically available, the pressures of practice are such that there are not unlimited resources in terms of time, instruments are limited, and so forth. Under these conditions rigour becomes a function of how well these forces, and the limitations that they impose on complete and precise knowledge, are managed. In this context precision is but one value among several – for example, competence and justice – that must be managed by practitioners in their quest for rigour (Argyris et al. 1985). An important aspect of this concept of rigour is that accuracy is associated with greater understanding of the whole, rather than by increasing precision of measurement in terms of increasingly smaller aspects of reality (Argyris, 1980).

According to Argyris and Schön (1974) professional effectiveness requires practitioners to not only become competent in taking action in their respective disciplines, but also to simultaneously reflect on their action in order to learn from it. By implication this also requires reflection on the theory that informs the action. In action science terms this process is referred to as double-loop learning (Argyris & Schön, 1974). An important feature of *double-loop learning* is that it is based on *internal criticism* aimed at self-monitoring and accountability. Such intentional reflection on action is regarded as being particularly appropriate for situations that are complex, ambiguous and

deeply important to the actors involved, but simultaneously require action to be taken (Argyris et al. 1985) – features which are also typical of the assessment situation (Gregory, 2004). When applied to the assessment situation, this implies that practitioners apply critical consciousness in the process of taking action, for example by designing tests *in* the practice situation *for* the knowledge created.

As the name suggests, action science is committed to the enactment of scientific principles within the action context. This is accomplished by adopting the principles of productive as opposed to defensive reasoning (Argyris, 1993). Productive reasoning contains several familiar features of normal science (Argyris et al. 1985), namely an emphasis on intersubjectively verifiable data, explicit inferences and causal reasoning, disconfirmable propositions, and public testing. When using productive reasoning, people make their logic explicit and subject it to public testing. The tests are designed to be independent of the logic of the actor, thereby avoiding self-referential reasoning (Argyris, 2000; 1993). When applied to the field of IOP assessment, the features of productive reasoning represent additional criteria (beyond statistical criteria) for the process followed by practitioners in their assessment of individuals. For example, key additional criteria in terms of which to monitor professional action include consistency (tests are interpreted consistently with their design), congruity (acting consistently with espoused values) and testability (as regards the judgments and conclusions arrived at).

Defensive reasoning, on the other hand, is characterised by tacit premises on which causal explanations rest, tacit inference processes by which people move from premises to their conclusions, the use of "soft" data and an appeal to self-referential logic – that is, using the same logic of the person producing the conclusions, to test them (Argyris, 1993). Under these conditions the genuine testing of views and opinions is inhibited or at best weak, leading to error-enhancing conditions. It is apparent that deference to professional judgment may not prevent defensive reasoning from taking place.

Central to the idea of facilitating rigorous tests in practice is the requirement that a community of inquiry be created in practice for this purpose. While the notion of testing through validation studies is acknowledged in the IOP literature, the importance of a community of inquiry engaged in the testing of knowledge claims in the practice setting is typically not emphasised. It is worth considering that the subjecting of knowledge claims to the public scrutiny of a community of inquiry is inherent to the practice of science (Mouton & Marais, 1990). Action science proposes that this principle be enacted in the practice setting (Friedman, 2001).

In addition to an emphasis on rigour-in-action, as it were, action science specifies requirements for the form knowledge should take if it is to be implemented within the constraints of the action context. These requirements are encapsulated in the principle of *actionability* and extend further than requirements for mere applicability. For example, the creation of knowledge in the form of a correlation coefficient may be applicable, but in that form it is not yet actionable as it does not inform actors what to do. Actionable knowledge therefore relates to the knowledge used to implement external validity (Argyris, 1993).

The requirements for actionable knowledge include, amongst others, that a specification be given of the actions required in order to realise intended purposes, that the behavioural mechanisms postulated to be in operation are specified, that the causal reasoning employed be robustly testable in a specific context, and that the constraints of the action context do not prevent its implementation. Implied in the latter is the idea that implementation should not violate ethical principles. When the implementation of knowledge requires the creation

of conditions that are counter to prevailing ethical principles and democratic values (for example, coercive conditions), it is not considered actionable, given the counterproductive consequences that tend to arise from it (Argyris, 1993; Argyris et al. 1985).

Perhaps most central to action science is the normative stance that the realisation of values (such as effectiveness and justice) is limited to the extent that unilaterally controlling, coercive actions (Model 1) are employed, and that these values are promoted by actions aimed at creating mutual learning (Model 2). Model 1 is triggered in difficult, complex and threatening situations and tends to be characterised by defensive reasoning, which in turn is counter-productive to the creation of conditions for producing valid knowledge. Model 2 is characterised by productive reasoning which promotes the creation of a psychological climate conducive to producing valid knowledge. One of the provocative findings of action science research has been that while Model 2 is widely espoused, Model 1 tends to be the dominant in-use model in the difficult and complex situations within organisational and professional contexts – situations where Model 2 is most needed (Argyris, 1993). In translating these ideas to the field of IOP assessment, awareness is required of the possible interference of Model 1 behavioural strategies, with their associated counter-productive consequences. Such awareness will enable practitioners to monitor the unintended consequences of their actions so that they can know when they are at risk of making errors. The implication is that practitioners consciously perform continuous reflective inquiry and ongoing monitoring of the impact of their actions as they create knowledge about the people they assess.

The implication for IOP assessment is that validity is to be evaluated not only in terms of the specific instruments (although this remains important given the inferences drawn from them), but in terms of the quality of the broader process of inquiry. In this way, validity is not conceived of as mainly being a technical concept. Rather, it denotes a dynamic process – one that is experienced, has psychological impact and is realised through action.

Drawing on action science ideas it is proposed that a conception of validity that seeks to integrate science and practice in the IOP assessment context, will reflect the interpenetration of three simultaneous sets of concerns, namely rigour, psychological adequacy and the actionability of knowledge produced. Validity-in-action, as it were, can therefore be defined as the extent to which the actions of practitioners and other users reflect the simultaneous integration of rigour, psychological adequacy and actionability when generating and implementing knowledge about the attributes of people in the workplace. These values incorporate concerns for accuracy, justice and competence respectively, and may also be seen as representing the design parameters for implementing validity and evaluating practice. Figure 1 below provides a graphic depiction of the interrelationship of the dimensions of validity as an action concept.

Rigour in the practice situation is enhanced by adopting strategies that include, amongst others, an emphasis on generating knowledge specific to the particular concrete case in a specific context and implementing the principles of productive reasoning (as described above). This involves the recognition that the knowledge created by practitioners, in effect, represents theories containing interconnected propositions about individuals that must be subjected to tests, as all theories are required to be. When this is done, the logic of criterion-related validity is extended to specific contexts in the practice setting. It was noted above that the creation of a community of inquiry in the practice setting enhances such testing.

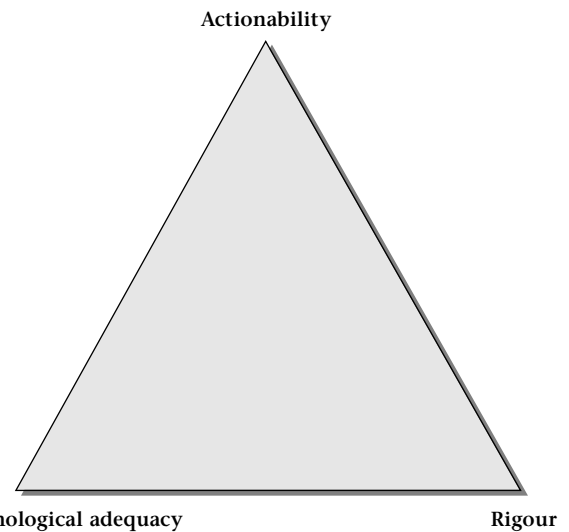


Figure 1: The dimensions of validity as an action concept

The establishment of an IOP assessment community of inquiry creates the opportunity for engaging in reflective conversation for purposes of error monitoring and initiating alternative lines of inquiry. Strictly speaking, the members of a community of inquiry should be included on the basis of their competence in a particular area, thus without infringing on one another's area of competence. It is expected that meetings between HR, the IO Psychologist and line management will take place as a matter of course. The proposition is that these parties should view it as an opportunity to deliberate as a community of inquiry, with the IO Psychologist assuming a leading role in facilitating productive reasoning. From the perspective of mutual learning and procedural justice, the test-taker also represents an indispensable member, implying that some degree of self insight is attributed, which qualifies him or her for participation. It stands to reason that the degree of self insight will differ, placing constraints on the testing of conclusions with the test-taker. It is, however, important that the practitioner tests the degree of self insight attributed to the individual, rather than acting on an assumption that the individual's participation in the community of inquiry is not possible. These requirements do not replace, but are in addition to the more familiar technical aspects of scientific rigour. The argument is that in view of the imperfect nature of instruments, high standards of rigour should be set for practice based on their use.

The term "psychological adequacy" is borrowed from Hosking and Morley (1991) to refer to the requirement that the practitioner act consistently with the empirical reality of dealing with individuals as valuable, unique, complex, adaptive and autonomous beings who shape and are shaped by their specific contexts, and are affected in important ways by attempts to subject them to scientific analysis. Of particular concern in this regard is the acknowledgement of the psychological consequences and ethical concerns related to the action of assessment.

That psychological assessment has a great impact on the lives of individuals is well accepted in the literature (Gregory, 2004; Taub, 2002; Sackett, Schmitt, Ellingson, & Kabin, 2001; Guba & Lincoln, 1989). It is clear that these concerns relate to actions taken in the process of assessment that are intimately related to issues of justice and perceptions of fairness. Several studies have drawn attention to the psychological impact of assessment procedures on the individual (Gilliland & Steiner, 2001; Francis-Smythe & Smith, 1997; Robertson, Iles, Gratton & Sharply, 1991). However, this aspect of validity – referred to as social or impact validity - has been recognised to be absent

from many selection procedures considered to be psychometrically valid (Gilliland & Steiner, 2001). Organisational justice researchers have identified procedural justice (the fairness and adequacy of the process) and interactional justice (a more subtle form of justice relating to the quality of interpersonal treatment by authority figures in their interpersonal communication) as increasingly important areas of focus in the field of personnel research and practice (Byrne & Cropanzano, 2001). Important criteria for demonstrating procedural justice have been identified as including: An opportunity to present one's views and give evidence of one's abilities; perceived control of the process and an opportunity to give input into the procedure; consistency of administration and an opportunity for reconsideration (Gilliland & Steiner, 2001). Interactional justice becomes relevant when authority has been ceded to others, with resulting concerns of exploitation and exclusion (for example, when granting permission for a psychological assessment to be conducted). Interactional justice is clearly relevant to the creation of conditions that promote valid information, which is unlikely to be produced under conditions of mistrust. These forms of justice are promoted by behavioural strategies aimed at increasing mutuality. Examples of how such strategies can be implemented in the IOP assessment situation include the following: Inviting candidates to participate as co-inquirers; acknowledging that instruments are not perfect and that they are used in the context of a broader inquiry; giving opportunity for findings to be disconfirmed within the context of a community of inquiry and making the constraints related to the candidate being placed in a dependent relationship both explicit and discussable (Argyris, 1976).

The criterion of actionability was discussed earlier and is not repeated here, apart from noting that its production places greater demands on the theoretical knowledge of practitioners, thereby forging a closer coupling of theory and practice.

Validity is enhanced when all three criteria are optimised. Conversely, to the extent that one of the criteria is not realised, it is undermined. For example, if rigour is deficient, errors are likely to be made leading to eventual ineffectiveness (reducing actionability) with non-trivial psychological impact. If psychological adequacy is not addressed, competence is undermined by virtue of unproductive behavioural dynamics that in the longer term may lead to ineffectiveness. The implication is that without monitoring the impact of an assessment, effectiveness may be undermined in ways that are more subtle and difficult to detect by virtue of counterproductive behavioural dynamics (for example lack of trust, resentment and so forth) that are likely to be set in motion. If actionability is not addressed, knowledge is unlikely to be experienced as useful or relevant, leading to the risk of users at the receiving end filling in perceived gaps in the advice they are getting on the basis of uninformed opinion.

It is proposed that the principles of mutual learning and productive reasoning in a spirit of continuous critical reflective inquiry are consistent with requirements for rigour, psychological adequacy and actionability, and that it is possible to realise these dimensions simultaneously by implementing these principles. Importantly, this requires a shift away from technical rationality to that of critical consciousness – a shift considered to be more psychologically adequate and appropriate to the essence of psychological assessment. In so doing it holds the potential of integrating science and practice to a greater extent than is currently the case.

While the emphasis is placed on broadening the concept of validity, it is important to note that this does not imply a

disregard for the technical aspects of assessment. The opposite may well be true as the competence of professionals is reflected in their skill at eliciting data from a variety of sources (including but not confined to, quantitative sources), making sense of it by drawing on their full repertoire of creativity, human sensitivity, theoretical understanding as well as their technical knowledge of a range of scientific techniques and instruments.

The implications of framing validity as an action concept include, amongst others, that the basis of a scientifically accountable model for practice is provided, that a fuller repertoire of skills on the part of the practitioner is released and that the validity spotlight is broadened to organisational users. It is further proposed that such a framing gives expression and allows for the integration of more recent approaches that emphasise its unitary nature as well as its consequential basis. At the same time the scientific adequacy of IOP assessment practices that do not take place within the context of a broader inquiry (for example through extensive use of canned computerised reports – even where criterion-related evidence is provided), is challenged.

On the basis of this conception of validity, it is proposed that the scientist-practitioner model of IO Psychology be strengthened by incorporating into the professional education of IOP assessment practitioners an awareness of the interpenetration of different epistemological levels when producing knowledge for the action context; by incorporating action science (particularly Model 2 behavioural skills) into professional education and cultivating critical-reflective skills that promote self-regulation within a community of reflective scientist-practitioners.

Perhaps most significant from the perspective of integrating theory and practice is that the framing of validity as an action concept requires and enables practitioners to act more congruently with the values central to their profession. This is analogous to a sentiment expressed by Snyman and Fasser (2004). In their analysis of the implications of postmodernism for ethics in psychology and psychotherapy, they write:

By taking responsibility for our thought systems, by questioning our presuppositions, and by acknowledging our epistemological positions, we further acknowledge the ethical basis of our therapies. Ethics and the ethical code of conduct in the healing professions are now more important than ever... psychotherapists are compelled to engage the field of ethics in a dynamic and personalised manner, so that the psychotherapist 'is the ethics' (2004, p.75).

Similarly, it can be said that *IOP practitioners are enabled to reflect validity in their actions* to the extent that the principles of rigour, actionability and psychological adequacy are enacted in their practice. In this sense it becomes possible for them to 'be validity', thus providing an image of the closest possible coupling of theory and practice.

## LIST OF REFERENCES

- Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7<sup>th</sup> ed.). Upper Saddle River, NJ.: Prentice-Hall.
- Argyris, C. (1976). Problems and new directions for industrial psychology. In M.D. Dunette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.
- Argyris, C. (1980). *The inner contradictions of rigorous research*. New York: Academic Press.
- Argyris, C. (1993). *Knowledge for action*. San Francisco, California: Jossey-Bass.

- Argyris, C. (2000). *Flawed advice and the management trap: How managers can know when they are getting good advice and when they are not*. New York: Oxford University Press.
- Argyris, C. & Schön, D.A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.
- Argyris, C. & Schön, D.A. (1996). *Organizational learning II: Theory, method and practice*. Reading, Mass.: Addison-Wesley Longman.
- Argyris, C., Putnam, R. & McLain Smith, D. (1985). *Action science*. San Francisco: Jossey-Bass.
- Austin, J.T. & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, 77 (6), 836-874.
- Braun, H.I. & Wainer, H. (1988). *Introduction*. In H.I. Braun & H. Wainer (Eds), *Test Validity*. Hillsdale, New Jersey: Erlbaum.
- Byrne, Z. S. & Cropanzano, R. (2001). The history of organizational justice: The founders speak. In R. Cropanzano (Ed.), *Justice in the workplace: From theory to practice* (Vol 2). London: Erlbaum.
- Campion, M.A., Outtz, J.L, Zedeck, S., Schmidt, F.L., Kehoe, J.F., Murphy, K.R. & Guion, R.M. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Journal of Personnel Psychology*, 54, 149-185.
- Cascio, W.F. (1995). Whither industrial and organizational psychology in a changing world of work? *American Psychologist*, 50 (11), 928-939.
- Cohen, R.J. & Swerdlik, M.E. (2002). *Psychological testing and assessment* (5<sup>th</sup> ed.). Boston: McGraw-Hill.
- Cronbach, L.J. (1988). *Five perspectives on the validity argument*. In H.I. Braun & H. Wainer (Eds), *Test validity*. Hillsdale, New Jersey: Erlbaum.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52 (4), 281-302.
- Dahlstrom, W.G. (1993). Tests: Small samples, large consequences. *American Psychologist*, 48 (4) 393-399.
- Denzin, N.K. (1994). The art and politics of interpretation. In N.K. Denzin & Y.S. Lincoln (Eds), *Handbook of Qualitative Research*. London: Sage.
- Dunnette, M.D. (1998). Emerging trends and vexing issues in industrial and organizational psychology. *Applied psychology: An International Review*, 47 (2), 129-153.
- Francis-Smythe, J. & Smith, P.M. (1997). The psychological impact of assessment in a development center. *Human Relations*, 50, 149-167.
- Friedman, V.J. (2001). Action science: Creating communities of inquiry in communities of practice. In P. Reason & H. Bradbury (Eds), *Handbook of action research: Participative inquiry and practice*. London: Sage.
- Gililand, S.W. & Steiner, D.D. (2001). Causes and consequences of applicant perceptions of unfairness. In R. Cropanzano (Ed.), *Justice in the workplace: From theory to practice* (Vol 2). London: Erlbaum.
- Gordon, G.B. (2001). Transforming lives: Towards bicultural competence. In P. Reason & H. Bradbury (Eds), *Handbook of action research: Participative inquiry and practice*. London: Sage.
- Gregory, R.J. (2004). *Psychological testing: History, principles, and applications* (4<sup>th</sup> ed.). Boston: Allyn & Bacon.
- Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guion, R.M. (1976). Recruiting, selection, and job placement. In M.D. Dunette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.
- Guion, R.M. (1980). On trinitarian doctrines of validity. *Professional Psychology*, 11 (3), 385-398.
- Guion, R.M. (1991). Personnel assessment, selection and placement. In M.D. Dunette & L.M. Hough (Eds), *Handbook of Industrial and Organisational Psychology* (2<sup>nd</sup> ed., Vol II). Palo Alto, Calif.: Consulting Psychologists.
- Guion, R.M. (1998). Some virtues of dissatisfaction in the science and practice of personnel selection. *Human Resource Management Review*. 8 (4), 351-365.
- Habermas, J. (1972). *Knowledge and human interests*. London: Heinemann.
- Health Professions Council of South Africa. (2001). *Policy on the classification of psychometric measuring devices, instruments, methods and techniques*. Pretoria: HPCSA.
- Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology*, 55, 363-396.
- Hosking, D. & Morley, I.E. (1991). *A social psychology of organizing: People, processes and contexts*. London: Harvester Wheatsheaf.
- Huysamen, G.K. (2002). The relevance of the new APA standards for educational and psychological testing for employment testing in South Africa. *South African Journal of Psychology*, 32 (2), 26-33.
- Kemmis, S. (2001). Exploring the relevance of critical theory for action research: Emancipatory action research in the footsteps of Jurgen Habermas. In P. Reason & H. Bradbury (Eds), *Handbook of action research: Participative inquiry and practice*. London: Sage.
- Kerlinger, F.N. (1978). *Behavioral research. A conceptual approach*. Fort Worth: Holt, Rinehart and Winston.
- Kerlinger, F.N. & Lee, H.B. (2000). *Foundations of behavioural research* (4<sup>th</sup> ed.). Australia: Wadsworth.
- Kipnis, D. (1987). Psychology and behavioural technology. *American Psychologist*, 42 (1), 30-36.
- Mattarazzo, J.D. (1990). Psychological assessment versus psychological testing. *American Psychologist*, 45 (9), 999-1017.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1988). *The once and future issues of validity: Assessing the meaning and consequences of measurement*. In H.I. Braun & H. Wainer (Eds), *Test validity*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Mouton, J. & Marais, H.C. (1990). *Basiese begrippe: Metodologie vir die geesteswetenskappe*. Pretoria: Gutenberg.
- Muchinsky, P.M. (2003). *Psychology applied to work*. Belmont, California: Thomson Wadsworth.
- Muchinsky, P.M. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples and recommendations. *Personnel Psychology*, 57 (1), 175-209.
- Murphy, K.R. & Davidshofer, C.O. (1994). *Psychological testing: Principles and applications* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Neuman, W.L. (2003). *Social research methods: Qualitative and quantitative approaches* (5<sup>th</sup> ed.). Boston: Allyn & Bacon.
- Reason, P. & Bradbury, H. (2001). Inquiry and participation in search of a world worthy of human aspiration. In P. Reason & H. Bradbury (Eds), *Handbook of action research: Participative inquiry and practice*. London: Sage.
- Robertson, I.T., Iles, P.A., Gratton, L. & Sharpley, D. (1991). The impact of personnel selection and assessment methods on candidates. *Human Relations*, 44 (9), 963-981.
- Romm, N. (1993). Habermas's theory of science. In J. Snyman (Ed.), *Conceptions of social inquiry*. Pretoria: HSRC.
- Sackett, P.R, Schmitt, N., Ellingson, J. & Kabin, M.B. (2001). High-stakes testing in employment, credentialing and higher education. *American Psychologist*, 56 (4), 302-318.
- Schmidt, C. (2006). *Integrating theory and practice in industrial and organisational psychological assessment - a meta-praxis perspective*. Doctoral dissertation, University of Johannesburg, Johannesburg.
- Schön, D.A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.



- Shulz, K.S., Riggs, M.L. & Kottke, J.L. (1998). The need for an evolving concept of validity in industrial and personnel psychology: Psychometric, legal and emerging issues (Electronic version). *Current psychology*, 17 (4). Retrieved November 26, 1999, from Academic Search Elite database.
- Snyman, J. (1993). Social science according to the Frankfurt School. In J. Snyman (Ed.). *Conceptions of social inquiry*. Pretoria: HSRC.
- Snyman, S. & Fasser, R. (2004). Thoughts on ethics, psychotherapy and postmodernism. *South African Journal of Psychology*, 34 (1), 72-83.
- Society for Industrial Psychology of South Africa (1998). *Guidelines for the validation and use of Assessment Procedures for the Workplace*. Johannesburg: Author.
- Taub, G.E. (2002). Moving beyond g: Linking theory, assessment and interpretation in the measurement of intelligence. *The International Journal of Sociology and Social Policy*, 22 (11), 132-149.
- Van den Berg , A.R. (1988). Oorwegings by die besluit om 'n spesifieke toets vir 'n gegewe doel te gebruik. In K. Owen, & J.J. Taljaard (Eds), *Handleiding vir die gebruik van sielkundige en skoolastiese toetse van IPEN en die NIPN*. Pretoria: Gutenberg.