Raven, J. (2008, August 20). Fundamental Problems in Psychometrics. *WebPsychEmpiricist*. Retrieved August 20, 2008 from http://wpe.info/papers_table.html.

WPE WebPsychEmpiricist

Fundamental Problems in Psychometrics

August 20, 2008

John Raven

Edinburgh

Abstract

Many of the national Professional associations that constitute the International Association of Applied Psychology and The International Test Commission in particular promote themselves as protecting the public from the misuse of tests. In this short paper it is argued that this claim is largely fraudulent. The Test Standards promoted by these bodies often lead to evaluations of both people and programmes that can only be considered unscientific, misleading, and unethical. The ethical issues become still more serious when those studies are presented as contributions to "evidence based practice". The paper is linked to an attempt to promote a paradigm shift in psychometrics currently being pursued through the Psychwiki.

An earlier version of this paper was published in the newsletter of the International Test Commission, 2008, Vol 19, p. 16-17. Address correspondence to: John Raven, 30 Great King St., Edinburgh EH3 6QH. Phone: (00 44) (0)1315562912. email: jraven@ednet.co.uk

Fundamental Problems in Psychometrics

August 20, 2008

The International Testing Commission (ITC) aims, among other things, to "promote responsible and valid tests and testing" (see www.intestcom.org).

Unfortunately, many widely accepted, indeed prescribed, methods and practices in testing cannot be regarded as anything other than unscientific and unethical.

The dilemma was highlighted by Spearman almost a century ago. He argued that the tests from which his ***g*** had emerged "had no place in schools" because they did not encourage teachers to identify and nurture the diverse talents of their pupils. To underline the point, he went on to assert that all pupils were geniuses at something but that this could not be demonstrated using current psychometric procedures (Spearman, 1924).

The evidence we have accumulated[1] suggests that he was right on all counts.

Failure to develop a more appropriate psychometric framework has even more serious consequences than failing to help parents, teachers, managers, and others to identify, develop, utilise, and reward the huge variety of talents that are available - thereby stunting most people's individual growth and depriving them of opportunities to gain recognition for their talents. The most serious consequence is that, because the neglected talents are the very ones that are required to transform our society in such a way that *homo sapiens* will have any chance of surviving as a species, continued reliance on the current testing framework contributes directly to our extinction … and probably that of most other species at the same time[2]. What could be more unethical?

The deleterious effects of this process in itself are exacerbated by the publication of numerous studies which, while purporting to contribute to "evidence based practice" in education and health care, are, in reality, incapable, not only of documenting the diverse ways in which people change as a result of involvement in developmental activities[3], but even the overall, desired and desirable, and undesired and undesirable, effects of the programmes evaluated[4].

An example may help to make the point.

Many of those involved in "progressive" education seek to nurture qualities like self-confidence, problem-solving ability, initiative, and the ability to understand and intervene in organisations and society. Furthermore, they try to help each of their pupils to develop their

idiosyncratic talents[5]. Since there are no good measures of such outcomes, most comparative evaluations utilise only traditional measures, mostly just of "the basics", such as reading. Since the "progressive" teachers did not set out to produce higher reading scores (at least as conventionally measured), their pupils do no better on these tests than pupils who have studied in other programmes. Politicians take this as a signal to close the programmes. Worse, the destructive effects of "traditional" education do not show up. The failure of these studies to document pupils' personal development (or deterioration) in a wide variety of *different* directions is a still more serious defect that there is not space to pursue here[6].

These problems could be ameliorated if the ITC *Standards* insisted that evaluations of both individuals and programmes be *comprehensive*. But, while such a move would be important, it would not be sufficient – because the *way* we have tried to "measure" individual differences is off beam.

To see this, let us substitute the word "creative" for "genius" in Spearman's claim. It would then read "Everyone is creative at something: The question is not '*How* creative are they?' but '*At what* are they creative?'".

Think about it. Is someone who is highly creative at causing disruption in his or her classroom or work organisation likely to display that creativity if a psychologist gives him or her a box of wooden blocks and asks them to "be as creative as possible"?

In fact, creativity, thinking, initiating "experimental interactions with the environment" and learning from the effects of those actions, persisting, and so on are all difficult and demanding activities that people will not display unless they are engaged in activities that are of great concern to them[7]. It follows that these qualities cannot be meaningfully "measured" unless one has first identified the kind of activity the individual is strongly predisposed to undertake and then created a situation in which one can investigate which talents they bring to bear whilst undertaking activities they care about.

Yet all of these talents, better termed components of competence or high-level executive functions, are crucial to effective action.

So, how to think about this situation?

An analogy may help.

Dogs, hawks, and whales all need hearts, brains, eyes, lungs, and blood to function.

But it would not make sense to try to base our main framework for differentiating between animals on variance in their heartiness, braininess, or quality of their perceptual system.

Nor would it make sense to rate all animals on "scales" "measuring" dogginess, hawkishness, whaleiness, or snakeiness.

What are the implications?

The analogy suggests that we first need a branching *descriptive* classification, or framework, similar to that used in biology to help us identify the *kind* of person we are dealing with … the kinds of things at which he or she is likely to be a genius (putting people at ease, creating political turbulence, pursuing adventurous research, etc.).

And then we need to determine *which* components of competence ("intuitively" grasping the situation, initiating "experimental interactions with the environment", learning from the effects of those actions, enlisting the help of other people, persisting etc) the individual brings to bear to undertake his or her "chosen" activities. (Perhaps, in a second stage, one might assess how good they are at doing each of these things *in the context of their chosen activity*.)

A subset[8] of the transformative processes that occur in some homes, schools, workplaces and adult developmental activities would then be understood as arising mainly from people finding themselves in environments that tap and harness their motives and lead them to utilise, develop, and display high level components of competence.

When this analogy is pursued, it becomes clear that the way we have sought to model and study the interactions between people and their environments has also been way off beam. For what is required is some kind of ecological mapping of the multiple feedback loops and interactions between people and their environments.

To underline the points that have been made in this brief article let us ask: "Where would biologists have got to if they had sought to summarise the variance between animals in terms of 1, 2, 5, or 16 "variables", the variance in their environments in terms of 10, and the interactions between the two sets of variables as a series of multiple regression weights?"

*****

The problems hinted at above were reviewed in a symposium entitled ***Serious Errors in the Evaluation of Individuals and Programmes arising from the use of tests yielding Arbitrary Metrics and from the deployment of Arbitrary selections of Measures*** at the ITC conference in

Liverpool July 2008. They are also reviewed in the PsychWiki as a basis for a "*Virtual Lab Meeting*" on ***Progressing a Paradigm Shift in Psychometrics***. All readers are encouraged to contribute to this. To do so, please click http://www.psychwiki.com/wiki/Progressing_a_Paradigm_Shift_in_Psychometrics

Notes

1. See, e.g. Raven (1994) and Raven, J., & Stephenson, J. (Eds.). (2001)
2. Raven, J. (2008)
3 Stephenson, J. (2001), Kazdin, A. (2006)
4. Raven, J. (1991)
5. Raven, J. (1994)
6. See Notes 3 and 8.
7. See several chapters in Raven, J., & Stephenson, J. (Eds.). (2001).
8. There is more that needs to be said about the seriously misleading – unethical – errors that have been made in the evaluation of transformative programmes in adult education, drugs based health care, and psychotherapy especially when these are presented as contributing to "evidence based treatment" and "payment by results", but there is not space here. Readers should turn, in particular, to the home page for the PsychWiki "virtual lab meeting" on *Progressing a Paradigm Shift in Psychometrics*.

References

Kazdin, A. E. (2006). Arbitrary metrics: Implications for identifying evidence-based treatments. *American Psychologist, 61,* 42-49.

Raven, J. (1991). *The Tragic Illusion: Educational Testing*. New York: Trillium Press. www.rfwp.com

Raven, J. (1994). *Managing Education for Effective Schooling: The Most Important Problem is to Come to Terms with Values*. Unionville, New York: Trillium Press. www.rfwp.com

Raven, J. (2008). *Intelligence, engineered invisibility, and the destruction of life on earth.* In J. Raven, & J. Raven, (Eds.), *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics*. Unionville, New York: Royal Fireworks Press; Edinburgh, Scotland: Competency Motivation Project; Budapest, Hungary: EDGE 2000; Cluj Napoca, Romania: Romanian Psychological Testing Services SRL.

Raven, J., & Stephenson, J. (Eds.). (2001). *Competence in the Learning Society*. New York: Peter Lang.

Spearman, C. (1924). Some issues in the theory of *g* (Including the law of diminishing returns). Proceedings of the British Association for the Advancement of Science: Section J-Psychology. 174-181. Southampton, England.

Stephenson, J. (2001). Inputs and outcomes: The experience of independent study at NELP. Chapter 21 *in* J. Raven & J. Stephenson (Eds.), *Competence in the Learning Society*. New York: Peter Lang.