

## How well are they? The relationship between wellbeing and assessment

Isabel Nisbet & Stuart Shaw

Wellbeing is increasingly cited as an objective of education – or as part of the school curriculum - and there is much discussion about wellbeing more widely, in public policy. In the world of educational assessment, attention has rightly been paid to the effect that assessments have on how people feel, particularly candidates, and the effect on performance and test validity (Putwain, 2009, Putwain & Daly, 2014). But less attention has been paid to the assessment of wellbeing: what is it, and can it – or should it – be assessed? We have turned our minds to those questions and in this note we share our conclusions so far, and we offer five propositions for the reader to consider.

Two models of wellbeing are prevalent within the literature:

(1) In the *eudaimonic* model (from Aristotle), *eudaimonia* signifies “flourishing”, “fulfilment”, or “living well”. This concept is linked to ideas of what is distinctive and valuable in human nature and/or what is worthwhile in life (White, 2007).

(2) The *hedonic* model relates to feelings/emotions, connoting a balance of positive feelings (e.g. pleasure, satisfaction, feeling good) over negative ones (feeling sad, anxious, worried).

Most surveys of wellbeing may appear to be using the hedonic concept by asking people how they feel. However, in practice we think they combine the two approaches – for example, in selecting what aspects of life to ask about, such as friendships, sporting and cultural pursuits and family life. These are, by implication, relevant to the views of the compilers of the review about what is worthwhile in life. Individuals may take different views about some aspects of “living well”. But in our view it is wrong to see wellbeing as just about measuring feelings. This leads to our first proposition:

**Proposition 1: Both models are important, but the eudaimonic element is essential.**

The philosopher W.B Gallie (1955/6) introduced the notion of an ‘essentially contested concept’. He was thinking of moral and aesthetic terms, where there may not be a permanent consensus on what they mean and on how to apply them, but they are nevertheless important and worth discussing. We think “wellbeing” is an essentially contested concept. That presents a problem for assessment, in the apparent absence of a clear theorised construct. This dilemma is expressed in our second proposition:

**Proposition 2: assessment needs to cover important ‘essentially contested’ concepts, including wellbeing.**

Several countries – Scotland, for example – identify health and wellbeing as a “subject” in the national curriculum, with content including understanding of nutrition and of the value of exercise. Some citizenship courses test knowledge of national

values (e.g. Australia has used test questions along the lines of “Is it acceptable to show disrespect towards people of Jewish faith?”). However, being able to describe right behaviour or goodness is not the same as behaving well or being a good person. Real moral debate, with participants encouraged to debate their beliefs with others and identify common or diverging underlying values, is a different thing again. This leads to our third proposition:

**Proposition 3: Understanding wellbeing and being able to discuss it with others are good – and can be tested – but are not the same as being well.**

One review of 99 self-report measures for assessing the wellbeing of adults identified no fewer than 196 dimensions of the constructs being measured (Linton, Dieppe & Medina-Lara, 2016). Examples of scales varying in their preference for eudaimonic and hedonic concepts include:

- The Warwick-Edinburgh scale
- The PANAS scale (Positive and Negative Affect Schedule) (Watson, Clark & Tellegen, 1988), and
- The Affectometer2 - developed from work in New Zealand (Tennant, Joseph & Stewart-Brown, 2007): an inventory of general happiness.

Other surveys purport to measure life satisfaction:

- The Satisfaction with Life Scale (SWLS) (Diener, Emmons, Larsen & Griffin, 1985).
- The ‘Cantril Ladder’ Self-Anchoring Striving Scale which ranks people’s lives relative to the best and worst possible lives, according to the respondents’ own objectives, values, and principles (Cantril, 1965; Kilpatrick & Cantril, 1960)

In medicine and social care models for assessing how well patients are often use a variety of sources. UK readers may be familiar with the “Ages and Stages” questionnaires used to screen infants and many countries use multidimensional models for assessing elderly patients. In contrast, most psychological scales for measuring wellbeing are based entirely on questions to the subjects - either about how they feel (or have felt) or what they think about their own life. On the one hand this seems common sense – who knows better than me how I feel? However, most educational assessments – for example, a mathematics test - do not rely solely on the test-taker’s self-perception.

There are technical risks for assessments based on responses by the subject, and these are familiar to academic psychologists who analyse the responses. They include the risk of “social acceptability bias” (see, for example, Paulhus, 1984) (that respondents will give answers to questions that they think will be expected or look good to others) and “self-deception bias” (Mijović-Prelec & Prelec, 2010) (when the self is perceived in an unrealistically favourable – or unfavourable – light). And there is a strong literature in psychiatry about the dangers of self-diagnosis of mental illness. In our view education can learn from these other disciplines, and that leads to our fourth proposition:

**Proposition 4: The subject is NOT the only source of evidence about how well they are.**

Who should set the assessment standards against which to compare respondents' perceptions of their wellbeing? What should these standards be? The Children's Society reported in 2022 that many young people were reporting serious concern about their appearance. It is quite possible that the ideals that the young people had in mind were not the standards that we would use. It is important to know how the young people were feeling, but there would be questions of validity of use if their perception alone prompted individual interventions.

On the whole, attempts to measure wellbeing through surveys or questionnaires have been seeking collective conclusions rather than conclusions about individuals. Assessing wellbeing at a *collective level* has clear public benefits: findings can inform the development and evaluation of policies and priorities for public services. However, the wellbeing of individuals, including learners in educational settings, depends on many factors, extending well beyond education. If a child is unhappy, that might reflect the child's experience of education; treatment she is receiving from her community nurse; the cost of living; her parents' marriage breakdown; or a combination of all these things (and other things).

Regarding wellbeing at the individual level, particularly if that individual is the only or primary source of information used for the assessment, technical risks of bias may even themselves out when larger numbers of responses are considered, but they could seriously distort an individual response. One key concern is about drawing conclusions for action – and particularly for professional intervention – on individuals based on their responses, unmoderated by wider observations. Some individuals could be particularly at risk of conclusions that failed to take account of their particular circumstances. In psychiatry there is a debate about the risk of bias in the norms used in the diagnosis of autism (Wilkenfeld & McCarthy, 2020). This raises the question for educationists of whether categories of “normal” used as a standard for assessing the wellbeing of all learners may overlook the legitimate interests and rights of divergent individual learners. We think that these risks merit the caution about uses of assessments of wellbeing that is reflected in our fifth proposition:

**Proposition 5: Surveying wellbeing collectively contributes to the public good; but it is ethically dangerous to draw conclusions leading to action at the individual level.**

**Conclusion** If you agree that wellbeing is an important educational objective, that raises the question of how – and whether – it can or should be assessed. Muddy

muddy concepts can be rich and important in the real world, even though “essentially contested” constructs are difficult to assess. In our view wellbeing deserves its recent prominence in thinking about education – and hence about assessment. Do you agree with our propositions? And are there wider lessons for assessment from this discussion? (1,370 words)

## References

- Cantril, H. (1965). *The Patterns of Human Concern*, New Brunswick, NJ: Rutgers University Press.
- Diener, E., Emmons, R., Larsen, R. J. & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*. 49(1):71-5.  
DOI: 10.1207/s15327752ipa4901 13.
- Gallie (1955/56). W.B. Gallie. Essentially contested concepts. *Proceedings of the Aristotelian Society*, 56 (1955/56), p.167.
- Kilpatrick, F. P., & Cantril, H. (1960). Self-Anchoring Scaling: A Measure of Individuals' Unique Reality Worlds. *Journal of Individual Psychology*, 16, 158-173.
- Linton M-J, Dieppe P, Medina-Lara A (2016). Review of 99 self-report measures for assessing well being in adults: exploring dimensions of well-being and developments over time. *BMJ Open* 2016;6:e010641. doi:10.1136/bmjopen-2015 010641.
- Mijović-Prelec D., & Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Transactions of the Royal Society B: Biological Sciences*. 365(1538), pp.27-40. doi: 10.1098/rstb.2009.0218.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 1984-03, Vol.46 (3), p.598-609.
- Putwain, D.W. (2009). Assessment and examination stress in Key Stage 4. *British Educational Research Journal*, 35(3), 391-411.
- Putwain, D.W. & Daly, A.L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies*, 40, 554-570).
- Tennant R., Hiller L., Fishwick R., Platt S., Joseph S., Weich S., Parkinson J., Secker J., Stewart-Brown S. (2007). The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): Development and UK Validation. *Health and Quality of Life Outcomes*. 5(63). doi: 10.1186/1477-7525-5-63.
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037//0022-3514.54. 6.1063.
- White, J. (2007). "Wellbeing and Education: Issues of Culture and Authority", *Journal of Philosophy of Education*, 41:1, 17-28. <https://doi.org/10.1111/j.1467-9752.2007.00540.x>.
- Wilkenfeld, D. A., & McCarthy, A.M. (2020). Ethical Concerns with Applied Behaviour Analysis for Autism Spectrum Disorder. *Kennedy Institute of Ethics Journal* 30 (1): 36-69. DOI: [10.1353/ken.2020.0000](https://doi.org/10.1353/ken.2020.0000).