Qual Quant (2010) 44:749–759
DOI 10.1007/s11135-009-9224-0

RESEARCH NOTE

Face validity in personality tests: psychometric instruments and projective techniques in comparison

Riccardo Sartori

Published online: 17 October 2009

Abstract Face validity is a controversial kind of test validity. Personality tests are divided into two big categories: projective techniques and psychometric instruments. They differ also for face validity, which influences the perception that people have of tests themselves. The article reports the scientific debate on face validity, and the results of a study carried out on naive subjects in order to let them compare projective techniques and psychometric instruments on the mere basis of their surface. An ad hoc questionnaire was administered. It asked subjects to compare projective techniques and psychometric instruments by using 13 adjectives. The sample, accidental, is composed of 238 participants, 45 males and 193 females. The data were substantially analyzed through techniques of Correspondence Analysis. Personality tests are principally judged through two dimensions: the aesthetic and the efficacy. The first dimension characterizes in particular projective techniques; the second, psychometric instruments. Although participants acknowledge that psychometric instruments are credible and scientific, there is a clear preference for projective techniques, principally by females, people younger than 22 and participants with lower education. Personality tests have an appearance that is judged by those who look at them. The aesthetic seems to prevail on the efficacy perception, but it would be suitable to carry on the same research with a sample stratified in respect of the personal details measured by the questionnaire.

Keywords Psychological testing · Personality tests · Psychometric instruments · Projective techniques · Face validity · Measurement

1 Introduction

The term validity, referred to psychological tests, indicates "the degree to which the test actually measures what it purports to measure" (Anastasi and Urbina 1997, p. 8). Although from

R. Sartori (✉)
Verona University, c/o Centro Docimologico, Via S. Francesco, 22, 37129 Verona, Italy
e-mail: riccardo.sartori@univr.it

a theoretical point of view validity represents a well defined characteristic of tests, authors generally agree with two facts (cf. Sartori and Pasini 2007):

1. From its birth to today, the concept of validity has changed and evolved (Borsboom et al. 2004): from the question of whether a test measures what it intends to measure (Kelley 1927; Cattell 1946), to the question of whether the empirical relations between test scores match some defined theoretical relations (Cronbach and Meehl 1955), to the question of whether interpretations and actions based on test scores are justified, not only scientifically, but also socially and ethically (Messick 1980, 1993, 1998);
2. There are at least two ways of considering validity, represented by two theoretical models: the *Unitary Concept of Validity* supports that validity is a unified characteristic of a test (cf. Messick 1995); the *Trinitarian Model of Validity*, instead, supports that validity must be estimated by monitoring three different aspects of it: content, criterion and construct.

Face validity is not contemplated, at present, in any of the two models, even if, in the origin, it used to belong to the second way of considering validity. Its scientific relevance and concrete importance are still today topic of discussion between those who sustain that it is a methodological mistake not to take care of this aspect anymore and those who think that it is instead misleading and even dangerous to number face validity among the criteria of test validity.

Personality tests are divided into two big categories: projective techniques and psychometric instruments. The first kind of tests is made of ambiguous stimuli as inkblots, pictures and incomplete sentences that subjects must interpret, describe and complete according to their personal way of feeling. The second kind of tests is made of a series of questions or statements (in one word: items), and subjects must say if and how much they describe their behaviour by using a category rating.

The two kinds of tests are totally different, generated by distant theoretical models, but they share the general aim for whom they were constructed: personality assessment. Both are not free from criticism, but projective techniques, more than psychometric instruments, are object of heavy accusation by the Scientific Community, for they are judged weak and ineffective (cf. Cornoldi and Tressoldi 2005; Sartori and Bortolani 2006). Among other problems, authors say that projective techniques lack face validity. Thus, they sometimes make subjects feel as a cavy in the hands of a wizard (Anastasi 1988; Kline 1993; Pedrabissi and Santinello 1997).

The article aims at pursuing the following two objectives:

1. Follow the stages of the scientific debate on face validity, in order to point out positive and negative aspects of it;
2. Report the results of a study carried out in order to operate a comparison between psychometric and projective instruments of personality inquiry from the point of view of the perception of the so-called naive people. In this sense, the research deals with face validity of personality tests.

2 Face validity

A test has face validity when it appears *valid* to examinees who take it, personnel who administer it and other untrained observers (Anastasi 1988). Most researchers have defined face validity as pertaining to a superficial examination of a test by non-experts. For instance,

(Cronbach, 1984, p. 182) indicates that "a test that seems relevant to the lay person is said to have 'face validity"'.

Face validity does not depend on established theories for support (Fink 1995). Cronbach (1971) himself believes that face validity is vague and subjective. Ingram (1977, p. 18) defines face validity as "surface credibility or public accountability". Alderson et al. (1995, p. 172) consider it "holistic rather than analytic".

The expression face validity simply indicates the operation of making a decision about the appropriateness of using some particular measurement instrument in a given assessment situation through the process of mere inspection of that instrument. But, even if a test has face validity, it does not mean that it is valid in the technical sense of the term. In fact, it does not exist any logical relation between face validity and construct validity: there are cases in which they are positively related and cases, as for example the personnel selection, in which high face validity represents only a disadvantage (Cattell and Warburton 1967).

For these reasons, the concept of face validity has become controversial. Some authors think it is useful (e.g., Roberts 2000). Some others think it is dangerous (e.g., Newfields 2002). Roberts (2000) claims that face validity is an essential part of the assessment process for these three reasons:

1. Since today there is a great deal of instruments that might be used for the measurement of the same construct, the professional, who only wants to use one of them, can narrow down the options of choice through the exploration of the components of tests and choose the more adequate instrument, given the specific context of administration;
2. The exploration of the test can also provide useful information about caveats that need to be adhered to, and/or ways in which the use of an instrument could be augmented, so as to make the overall assessment more meaningful to those who administer it, decision makers, students, teachers, and so on;
3. In general, a good face validity investigation can provide, in cases where decisions have to be made quickly and without time to do more thorough documentation, some "*evidence*" as to whether the instrument will fulfil the intended purpose or not.

Although these positive aspects of face validity, there are some voices of dissent. Hajipournezhad (2003) mentions how the expression "face validity" is widely detested among testing scholars. Quoting Mosier (1947, p. 194) he emphasizes: "The concept is the more dangerous because it is glib and comforting to those whose lack of time, resources, or competence prevent them from demonstrating validity (or invalidity) by any other method […] This notion is also gratifying to the ego of the unwary test constructor". Trochim (2002) cautions that face validity is the weakest way to try to demonstrate construct validity. Lacity and Jansen (1994) describe face validity in terms of persuasive appeal and note that test items can seem persuasive even if they lack internal validity.

On this subject, Newfields (2002) writes that:

1. Face validity is a contradictory term. Matters involving surface appearance can have cosmetic value, but not validity. Validity should involve deeper factors, such as logical veracity, consistency, and congruence;
2. If we regard testing as a rigorous discipline, face validity has little place because it is both atheoretical and usually imprecise. Face validity basically amounts to what Buck (2001) refers to as *faith validity* (the belief that a test is OK without empirical evidence). Empirical evidence is a sine qua non of testing. Since face validity is based primarily on the judgements of novices, this concept might be interesting in terms of a business marketing perspective, but it is not a validation yardstick test developers should focus on.

In addition to this, Kline (1993, 1998) notes that high face validity coincides with a high probability that subjects guess what tests try to measure. This makes subject lie and distort their responses, especially in case of personnel selection. This has undesirable effects, for example in case of inquiry on personality traits.

Face validity is generally considered to present this kind of problems:

1. Face validity is product-oriented. That is, the informants, whether be experts or laymen, consider the test only after it has been constructed. They have no chance to get familiarized with the assumptions underlying the specific test construction;
2. Face validity claims to come up with "valid reports" by informants, that is to say, the ones who have been consulted often have a tendency to report the test as valid;
3. Face validity efforts come up with qualitative reports of whether the test represents what the given constructs demand. In other words, quantitative face validity, which uses rating scales and planned interviews, is rarely practiced;
4. Face validity, as normally practiced, involves superficial examination of test content. This leaves room for no in-depth analysis of test content compared with test constructs;
5. Face validity is a kind of judgment call which does not follow the principles of judgment surveys in general. Test developers usually ask one or two or very few observers to face validate the test. However, when conducting most face validity studies, often no appropriate statistical procedure is applied to the data to make the subsequent decisions statistically meaningful and dependable.
6. The sixth point deals with the tendency of many test constructors to use face validity as sufficient grounds for validity, obviating the need for other more quantitative measures of validity.

However, Roberts (2000) emphasizes that we do not have to follow this kind of conceptualization of face validity. He argues that we can use face validity with expert informants and not as a superficial examination. This, he believes, will turn face validity into a more effective measure, given that it would become more similar to the concept of *content validity*. It can be thought to refer to the degree that respondents or users judge that the items of a test are appropriate to the targeted construct and assessment objectives (Allen and Yen 1979; Nevo 1985). So, it can be used in order to measure the acceptability of the test to users and administrators. Furthermore, there are scholars supporting the application of face validity as it caters to response validity by enhancing applicants' acceptance of the testing procedure (Alderson et al. 1995, p. 173; Davies et al. 1999, p. 59). For example, face validity has gained a new status in the framework of communicative language testing which favours this type of validity due to its "real-life" definition of language proficiency (Carroll 1985).

Despite problems and caveats, face validity keeps its own utility, that is to say, it can avoid that subjects administered tests wonder such questions as: What on earth is this item for? When we use or construct a test we should ask ourselves if each item has or has not face validity. Items which have not face validity can appear strange to subjects and irritate them. As an extreme example we can think about a *question* and an *inkblot*: It is different to ask subjects how many times a day they wash their hands rather than to show them an inkblot in order to decide whether they are or not obsessive. The point is that a question is more likely to have face validity than an inkblot. And among questions there are some which have more face validity than others.

We do not know of researches carried out in order to test the hypothesis that instruments with higher face validity are really preferred by subjects. The aim of the study reported is

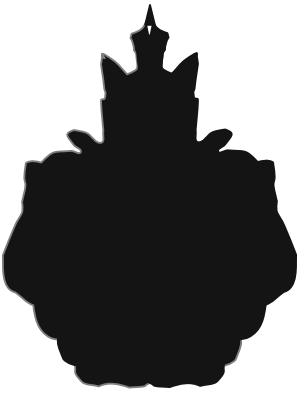
the following: compare projective techniques and psychometric instruments for personality inquiry on the basis of the perception of the so-called naive persons.

3 The research

3.1 Instrument and procedure

The instrument is an ad hoc questionnaire, divided into four parts:

1. The five items of the first part have the function of collecting the following personal data of the participants: sex, age, province of residence, last title of study, and principal occupation;
2. the second part of the questionnaire starts with the presentation of the following picture:



Type A

I like studying
❑ true ❑ false

I feel happy only when I am with someone else
❑ true ❑ false

I accept help from other people
❑ true ❑ false

I try to be honest in every situation
❑ true ❑ false

Type B

accompanied by the following description:
Imagine that the picture "Type A" and the picture "Type B" show each a part of a personality test, one of A-kind and one of B-kind. Of course the picture "Type A" does not show all the test (which can be thought to be composed by a larger number of inkblots of that kind), and the picture "Type B" just reports some of the statements which can compose the hypothetic personality test. People are requested to say what the inkblot could be in the case of personality tests of A-kind and how much they agree or do not agree with the statements in personality tests of B-kind. It is possible to imagine that the rating scale of the statements could be different from the one shown in picture "Type B" and have a larger number of levels. Both kinds of test result in a personality description of subjects.
The participants are thus put in the following condition:
Given that the picture "Type A" show a certain kind of personality test and the picture "Type B" another kind of personality test, please answer the following questions:
The first question asks:
Which of the two personality tests does seem to you more. . .:
13 adjectives follow:

. . . valid (able to test your personality)?
. . . beautiful (desirable)?
. . . credible (in respect of the profile it gives)?
. . . transparent (it does not hide its aims)?

... repeatable (to the same subjects twice or more times)?
... liable to falsification?
... scientific?
... mysterious?
... banal?
... interesting?
... informative (= revealing personality characteristics of subjects)?
... attractive?
... effective?

Each accompanied by the same response tetracoric scale:

☐ Type A
☐ Type B
☐ Both
☐ None

3. The third part of the questionnaire starts with the following question:
Which of the two kinds of personality tests would you prefer to be administered?
Accompanied by the same response scale:

☐ Type A
☐ Type B
☐ Both
☐ None

On the basis of the kind of response, subjects are invited to give one or more reasons, which can be chosen from the following:
seven if subjects answer Type A or Both:

☐ I am attracted by everything esoteric or mysterious
☐ I like the idea I can know what I am really like just by describing an inkblot
☐ I am interested in deep and unconscious aspects of personality
☐ I believe that only this kind of tests can tell something new about my personality
☐ I think it is impossible to cheat at a personality test of this kind
☐ I am curious to see what comes out
☐ I am curious to understand how my responses to a test of this kind can reveal my personality

seven if subjects answer Type B or Both:

☐ I only believe in what is clear and understandable
☐ I believe that I can know what I am really like only through the analysis of the behaviours represented by the statements of the test
☐ I am interested in what I am like and how I behave in my everyday life, not what my unconscious contains
☐ I believe that only this kind of test can tell something new about my personality
☐ I believe that this kind of test is more scientific than the other
☐ I am curious to see what comes out
☐ I am curious to understand how my responses to a test of this kind can reveal my personality

Table 1 Frequencies of males and females, students and workers, and descriptive statistics of the variable age for the total sample, the group of males and the group of females

				Age						
	N	Students	Workers	Min.	Max.	Mode	Median	Mean	S.d.	St. err.
Total sample	238	198	40	18	71	21	22	22.00	8.82	0.57
Males	45	22	23	19	71	22	27	21.80	12.36	1.84
Females	193	176	17	18	61	21	21	23.81	6.99	0.50

Table 2 Distribution of frequency of the variable "last title of study"

	Frequency	Percentage
Secondary School License	1	0.4
Professional School Diploma	3	1.3
School-Leaving Certificate	187	78.5
University Degree	39	16.4
Beyond (Ph.D. or Master)	8	3.4
Total	238	100

five if subjects answer None:

- ☐ I absolutely do not believe in personality tests
- ☐ I have some doubts about personality tests
- ☐ I believe that no test can tell things about my personality which are not known to me
- ☐ I am not interested to know myself more deeply
- ☐ I have some fears about personality tests

4. In the hypothesis that believing or not in Astrology can influence the preference for a kind of test or another, the fourth and last part of the questionnaire asks subjects the following question:

Do you believe that Astrology which is based on year, month, day and hour of birth can be able to reveal aspects of your character, temperament, personality and way of behaving?

4 Participants

The sample, accidental, is composed by 238 subjects, 45 males (18.9%) and 193 females (81.1%). Table 1 shows the frequencies of males and females, students and workers, and the descriptive statistics of the variable age for the total sample, the group of males and the group of females.

Participants especially come from the Italian Regions of Lombardia ($n = 120$, 50.4%) and Veneto ($n = 98$, 41.2%). The other twenty subjects live in (from North to South): Piemonte, Trentino, Friuli, Emilia Romagna, Toscana, Marche, Puglia and Sicilia.

The variable "last title of study" is distributed according to Table 2:

Table 3 Percentage of choice of the four alternatives (columns) for each of the 13 adjectives (rows), in the sample of 238 subjects (bold values are the modal percentage for each adjective)

Adjective	Type A (projective) (%)	Type B (psychometric) (%)	Both (%)	None (%)
Valid	18.9	36.6	**40.3**	4.2
Beautiful	**62.6**	20.6	8.0	8.8
Credible	11.3	**50.0**	29.8	8.8
Transparent	14.7	**71.4**	7.1	6.7
Repeatable	22.3	**37.0**	28.2	12.6
Liable to falsification	10.1	**52.1**	32.8	5.0
Scientific	8.0	**47.1**	26.5	18.5
Mysterious	**90.3**	1.3	2.1	6.3
Banal	4.6	31.5	6.3	**57.6**
Interesting	**60.1**	5.0	31.1	3.8
Informative	21.0	26.1	**47.9**	5.0
Attractive	**79.4**	3.4	10.9	6.3
Effective	12.2	25.2	**51.7**	10.9

5 Results

Table 3 shows the percentage of choice of the four alternatives (columns) for each of the 13 adjectives (rows), in the sample of 238 subjects.

The Table 3 shows that:

1. Tests of projective kind are described especially as beautiful, mysterious, interesting and attractive;
2. Tests of psychometric kind are especially described as credible, transparent, repeatable, liable to falsification and scientific;
3. Both kinds of tests are especially described as valid, informative and effective;
4. Finally, none is described as banal.

Correspondence Analysis (Optimal Scaling, Analysis of Homogeneities by Homals) and Hierarchical Cluster Analysis for categorical variables show the existence of two dimensions of assessment underlying the 13 adjectives. Also the Analysis of Principal Components leads to the same result.

The first dimension, named "efficacy", put together the adjectives: valid, credible, transparent, repeatable, liable to falsification, scientific, informative ed effective. This is the dimension that characterizes, in particular, psychometric tests.

The second dimension, named "aesthetic", put together the adjectives: beautiful, mysterious, interesting and attractive. This is the dimension that characterizes, in particular, projective techniques.

To the question: *Which of the two kinds of personality tests would you prefer to be administered?*, subjects have answered as shown in Table 4:

Considering only the answers Type A and Type B, and crossing this item with the variable sex, the result is that males ($n = 33$) show the tendency to prefer tests of psychometric kind, while females ($n = 129$) are more oriented towards tests of projective kind ($\chi^2 = 5.863$, $\mathrm{df} = 1$, $p = 0.015$). No tendency, instead, in those who believe in Astrology ($n = 56$) to prefer tests of projective kind and in those who do not believe ($n = 106$) to orientate towards

Table 4 Distribution of frequency of the item: *Which of the two kinds of personality test would you prefer to be administered?*

	Males		Females		Total	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Type A	20	44.4	104	53.9	124	52.1
Type B	13	28.9	25	13.0	38	16.0
Both	8	17.8	61	31.6	69	29.0
None	4	8.9	3	1.6	7	2.9
Total	45	100	193	100	238	100

tests of psychometric kind, even if the variable sex tends to be in relation with the answers given to the last item of the questionnaire ($\chi^2 = 3.676$, df $= 1$, $p = 0.055$). This means that females, more than males, declare they believe in Astrology (37.3% out of females versus 22.2% out of males) and show to prefer tests of projective kind. Also, younger subjects (with an age inferior to median) tend to prefer projective techniques, while older subjects (with an age superior to median) tend to prefer psychometric tests ($\chi^2 = 3.845$, df $= 1$, $p = 0.049$). In the same way, subjects with lower education (until the School-Leaving Certificate, $n = 191$) prefer projective techniques, while subjects with higher education (University Degree and Beyond, $n = 47$) prefer tests of psychometric kind ($\chi^2 = 9.918$, gdl $= 1$, $p = 0.002$).

6 Conclusions and discussion

Personality tests have their own appearance that is judged by those who look at them. According to the results of this research, the principal dimensions on which personality tests are judged are those of the aesthetic and the efficacy. Projective techniques *seem* more beautiful, mysterious, interesting and attractive (aesthetic assessment), while psychometric tests *seem* more credible, transparent, repeatable, liable to falsification and scientific (efficacy assessment). Although participants seem to acknowledge credibility and scientific nature to psychometric tests, projective techniques are clearly preferred, especially by females, people younger than 22 and subjects with lower education, maybe because they are perceived as less liable to falsification and more functional. In this sense, it seems to be not true the assertion according to which tests with higher face validity are better just in virtue of the fact that they seem to be less ambiguous and more clear.

Those who declare they prefer projective techniques ($n = 124$), while answering item 2, they say they prefer them for the following four reasons:

1. I am curious to understand how my responses to a test of this kind can reveal my personality (58.1%);
2. I am interested in deep and unconscious aspects of personality (50%);
3. I like the idea I can know what I am really like just by describing an inkblot (42.7%)
4. I am curious to see what comes out (40.3%);

Those who declare they prefer psychometric tests ($n = 38$), they say they prefer them for the following four reasons:

1. I believe that this kind of tests is more scientific than the other (42.1%);

2. I am curious to understand how my responses to a test of this kind can reveal my personality (36.8%);
3. I believe that I can know what I am really like only through the analysis of the behaviours represented by the statements of the test (29%);
4. I only believe in what is clear and understandable (28%).

Those who finally answer "None" ($n = 7$), they do that for the following three reasons:

1. I absolutely do not believe in personality tests;
2. I have some doubts about personality tests;
3. I have some fears about personality tests.

Note that, despite of the scarce number of people who do not want to be administered any personality test, there is a great deal ($n = 69$) who would be administered both. That is to say that the curiosity of knowing ourselves seems to be greater than the fear we prove about these instruments.

In conclusion, we report what Sartori writes in his article of Sartori (2003) entitled *The assessment of personality through tests* (pp. 160–163).

"As regards real efficacy and utility of personality tests, laymen are generally a little perplexed [...] Often this perplexity matches a certain kind of curiosity about what a test can reveal of us, especially if it is a personality test [...] People who do not believe much in personality tests, but are curious to know themselves or to see if personality tests really work, would prefer to be administered personality tests of projective kind [...] rather than personality tests of psychometric kind [...] This because people who are strongly critical and disparaging towards personality tests in general are also convinced that they can cheat very well at personality tests of psychometric kind [...] Vice versa these people would be more inclined to be administered personality tests which are more cryptic, as for example the projective ones (perhaps only to say that doing them was ineffective), are more attracted by the idea that describing inkblots can come out what they are really like and are also convinced that it is impossible to cheat at them. People who instead hate everything esoteric and mysterious would prefer to be administered personality tests of psychometric kind because these are judged as 'more scientific' [...] Of course there are people who would not be administered personality tests of one kind nor the other. They are in fact convinced that personality tests cannot add information to what they already know about themselves or, even, they think that tests are completely wrong in describing human personality and do not absolutely work".

References

Alderson, J.C., Clapham, C., Wall, D.: Language Test Construction and Evaluation. Cambridge University Press, Cambridge (1995)

Allen, M., Yen, W.M.: Introduction to Measurement Theory. Brooks/Cole, Monterey, CA (1979)

Anastasi, A.: Psychological Testing. Macmillan, New York, NY (1988)

Anastasi, A., Urbina, S.: Psychological testing, 7th edn. Prentice Hall, Upper Saddle River, NJ (1997)

Borsboom, D., Mellenbergh, G.J., van Heerden, J.: The concept of validity. Psychol. Rev. **111**(4), 1061–1071 (2004)

Buck, G.: Validities. Message posted on L-TESL. Online Forum. http://f05n16.cac.psu.edu/archives/ltest-l.html

Carroll, B.J.: Second language performance testing for university and professional contexts. In: Hauptman, P.C., LeBlanc, R., Wesche, M.B. (eds.) Second Language Performance Testing, University of Ottawa Press, Ottawa (1985)

Cattell, R.B.: Description and Measurement of Personality. World Book Company, New York (1946)

Cattell, R.B., Warburton, F.W.: Objective Personality and Motivation Tests. University of Illinois Press, Urbana (1967)

Cornoldi, C., Tressoldi, P.: Sulla scientificità delle tecniche proiettive: a proposito degli interventi apparsi su "Psicologia Clinica dello Sviluppo". Giornale Italiano di Psicologia **2**, 237–244 (2005)
Cronbach, L.J.: Validity. In: Thorndike, R.L. (ed.) Educational Measurement, 2nd edn, pp. 443–597. American Council on Education, Washington, DC (1971)
Cronbach, L.J.: Essentials of Psychological Testing. Harper and Row, New York (1984)
Cronbach, L.J., Meehl, P.E.: Construct validity in psychological tests. Psychol. Bull. **52**, 281–302 (1955)
Davies, A., Brown, A., Elder, C.: Dictionary of Language Testing. Cambridge University Press, Cambridge (1999)
Fink, A.: How to Ask Survey Questions. Sage, Thousand Oaks, CA (1995)
Hajipournezhad, G.: An Approach to the Validation of Judgments in Language Testing. http://www.jalt.org/pansig/2003/HTML/HajiPourNezhad.htm (2003)
Ingram, E.: Basic concepts in testing. In: Allen, J.P.B., Davies, A. (eds.) Testing and Experimental Methods, Oxford University Press, London (1977)
Kelley, T.L.: Interpretation of Educational Measurement. Macmillian, New York (1927)
Kline, P.: The Handbook of Psychological Testing. Routledge, London (1993)
Kline, P.: The New Psychometrics. Routledge, London (1998)
Lacity, M., Jansen, M.A.: Understanding qualitative data: A framework of text analysis methods. J. Manage. Inform. Syst. **11**, 137–166 (1994)
Messick, S.: Test validity and the ethics of measurement. Am. Psychol. **30**, 955–966 (1980)
Messick, S.: Validity. In: Linn, R.L. Educational Measurement, 2nd edn, pp. 13–104. American Council on Education and Oryx Press, Phoenix, AZ (1993)
Messick, S.: Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am. Psychol. **50**, 741–749 (1995)
Messick, S.: Test validity: A matter of consequence. Soc. Indic. Res. **45**, 35–44 (1998)
Mosier, C.I.: A critical examination of the concepts of face validity. Educ. Psychol. Meas. **7**, 191–205 (1947)
Nevo, B.: Face validity revisited. J. Educ. Meas. **22**, 287–293 (1985)
Newfields, T.: Challenging the notion of face validity. SHIKEN: JALT Test. Eval. SIG Newsl. **6**(3), 19. http://www.jalt.org/test/new_2.htm (2002)
Pedrabissi, L., Santinello, M.: I test psicologici. Il Mulino, Bologna (1997)
Roberts, D.M.: Face validity: Is there a place for this in measurement? SHIKEN: JALT Test. Eval. SIG Newsl. **4**(2), 5. http://www.jalt.org/test/rob_1.htm (2000)
Sartori, R.: La valutazione della personalità tramite test. Quaderni Dipav **6**, 159–178 (2003)
Sartori, R., Bortolani, E.: Approccio idiografico e approccio nomotetico alla persona: il caso dei test dei psicologici. Giornale Italiano di Psicologia **1**, 107–118 (2006)
Sartori, R., Pasini, M.: Quality and quantity in test validity: How can we be sure that psychological tests measure what they have to. Qual. Quant. **41**(3), 359–374 (2007)
Trochim, W.: Measurement validity types. http://trochim.omni.cornell.edu/kb/ (2002). Accessed 17 Apr 2002