URN: urn:nbn:de:bvb:473-irb-472419
DOI: https://doi.org/10.20378/irb-47241

Psychological Tests from a (fuzzy-)logical point of view

Abstract
Psychometric theory relies on two basic assumptions: a) psychological constructs refer to latent (unobservable) variables and b) psychological tests serve as a way to measure these constructs. This view is complemented by an alternative interpretation of psychological constructs, which neither relies on latent variables nor on the concept of measurement.

Using the formal apparatus of Many-Valued Logic, psychological constructs are re-interpreted as linguistic concepts (rather than latent variables), which can be inferred by means of logical calculus (as opposed to measurement). Thus, test scores do not refer to the values of latent variables, but to the degree to which the necessary and sufficient conditions for the ascription of a construct are fulfilled.

Following this rationale, a formal theory of psychological tests is developed, which models the process of testing as logical inference. Applying the derived procedures, a person's testing behaviour yields the degree to which a construct describes her adequately.

Keywords:
Psychometrics, Psychological Constructs, Psychological Tests, Latent Variables, Many-Valued Logic, Fuzzy Logic

1 Introduction
Psychological tests play a crucial role for both psychological research and applied psychology. They form the empirical basis for a great variety of psychological theories and provide a well-established tool for psychological diagnostics. The assessment of so called ‘constructs’ by psychological tests does not only determine the scope of ongoing research but also influences important decisions on an individual level. Despite the apparent consensus about the importance and usefulness of psychometric procedures, their theoretical basis has been a point of debate ever since (cf. Michell, 1999; Borsboom, 2005). At the core of this discussion stands the notion of the *psychological construct* (MacCorquodale and Meehl, 1948; Ginsberg, 1954; Slaney and Racine, 2013). Several attempts were made to bridge the ambiguities between a consistent theoretical explication of constructs and their assessment via psychometric tests. These range from a fundamental measurement perspective (Michell, 1990, 1999, 2013) over reflective test models (Edwards and Bagozzi, 2000; Markus and Borsboom, 2013) to the ‘deconstruction of constructs’ using a network approach (Cramer, Waldorp, Van der Maas and Borsboom, 2010; Borsboom and Cramer, 2013; Schmittmann et al., 2013).

Despite a considerable interest in psychometric (meta-)theory and it’s potential for the clarification of the meaning of test scores, to our knowledge there have been no serious attempts to model the relational structure between a construct and its meaning by means of formal semantics. Instead, nearly all psychometric approaches invoke some kind of measurement in order to quantify psychological constructs (cf. Kline, 1998). As a result, so far there is no formal theory of psychological tests, which deals with the inference of constructs on the basis of their underlying semantics.

The approach outlined in this paper tries to fill this theoretical gap, putting forward a formal account of psychological tests as a way to quantify constructs, without invoking such concepts as measurement, causality or latent variables. Testing is thereby re-interpreted as a way of logical inference, rather than a measurement procedure. Making extensive use of many-valued logical calculus, the approach yields a completely new way to justify quantified statements about constructs, allocating quantity not in the construct itself but in the truth-value of a logical expression.

2 Psychological Constructs as vague Concepts

Psychological constructs are often identified with latent variables, which are not directly observable (Borsboom, 2008). This view is consistent with the psychometric approach to measuring them using observable (manifest) variables. Psychological tests are thus conceived to yield indirect measures of unobservable, yet real psychological attributes (Borsboom, Mellenbergh and Van Heerden, 2003). This conceptualization of psychometric measurement finds a formal counterpart in the framework of Item Response Theory (IRT), which models test behavior as a (probabilistic) function of unobservable variables. The resulting model parameters are thereafter interpreted as quantitative measures of the underlying construct (Embretson and Reise, 2000).

However, this view is not uncontroversial. First of all, some authors doubt the common claim that IRT is a sound implementation of the Representational Theory of Measurement (e.g. Michell, 1999; Kyngdon, 2008a, 2008b). Secondly, there are strong philosophical reasons to question the identification of psychological constructs with latent variables (Maraun and Gabriel, 2013). This line of reasoning leads to an alternative view to psychological constructs, which takes them to be *linguistic concepts*, rather than real entities. According to this view, constructs are *abstractions* of observable behavior, not underlying causal factors (Slaney and Racine, 2013).

Although test results are sometimes interpreted in a linguistic way rather than in the sense of quantitative measures of psychological traits (e.g. Mirels, 1982; Borkenau, 1988), psychometric *methods* rely almost entirely on models of measurement. Despite its consistence and intuitive appeal, the concept view of psychological constructs has so far not been formalized to a degree which could justify a concept interpretation of existing psychometric models, let alone be used for new psychometric developements. To account for these issues, the following sections aim at developing a formal theory of psychological tests, which grounds on an interpretation of psychological constructs as linguistic concepts.

2.1 Reference vs. Meaning

The most distinctive feature of the concept view of psychological constructs is that the relation between a construct and observed behavior is not one of causality, but one of definition. It is, therefore, a *semantic* relation. The construct is to be understood as an umbrella term for the content of a test (i.e. the meaning of the test items). This, of course, is not to be confused with an operationalist view of psychological constructs, which would define a construct by its method of assessment. Rather, the definition of a construct is based on semantic grounds. It is the meaning of the test items that constitute a construct, not the way a test is used to produce numerical indices.

Given this, it follows that the meaning of a construct can be fully determined by its constituting test items. It is therefore possible to talk meaningfully about a construct, without discussing its referent. Whether or not a construct refers to a real entity or not, does not lie within the scope of defining it. Neither does it bare any relevance to the question how an abstract term (like a psychological construct) is to be used in a semantically correct way. If constructs serve to subsume a collection of observables, there is no need to invoke latent variables or related concepts in order to define a construct. Thus, when we speak of a construct meaning we refer to its semantic relation to concrete terms, marking the question of causal theory as irrelevant for our purposes.

2.2 Error of Measurement vs. Semantic Vagueness

A fundamental problem regarding the definition of psychological constructs is the fact, that most constructs are inherently *vague*. Hardly any construct can be defined in an unambiguous way – it seems to be nearly impossible to give sharp boundaries for the meaning of most psychological constructs. Although the issue of semantic vagueness has long been recognized in Psychology (Blumer, 1940), it has not been incorporated in psychometric theory so far. Instead of modelling semantic vagueness of construct meanings, the received way to deal with imprecision is to subsume it under the label of 'measurement error'. However, in some cases this may lead to implausible or even absurd implications, as illustrated in the following example (cf. Buntins, 2014).

Let the following items be part of a test for the assessment of the construct ‘Extraversion’:

a) I like to meet new people.
b) I like to go to parties.

Now imagine Peter highly agrees with item a) but does not agree with item b). Is it reasonable to conceptualize this deviation as a form of measurement error? If so, what exactly is meant by ‘error’ in this case? Did Peter *accidentally* give wrong answers? Or maybe he does not know himself, with which items he agrees? We have to conclude that even if the concept of ‘random error’ is a convenient way to model response variation in psychological tests, it is often difficult to conceive the ‘randomness’ of the underlying behavior.

A more plausible interpretation of the above scenario would be to conclude that *Peter likes to go to parties for different reasons than meeting new people*. Note that there is no randomness involved in the fact that people go to parties for different reasons.

Instead of modelling error of measurement, the account put forward in this paper follows a different rationale*: within subject response variation in psychological tests is essentially the result of vaguely defined constructs*. Thus, the main target of the following sections lies in giving a sound account of semantic vagueness in the context of psychological tests.

3 Modelling semantic Vagueness

The following sections try to develop a formal theory of psychological tests which does not rely on traditional concepts like latent variables, measurement or random error. Instead, psychological constructs are modeled as semantically vague linguistic concepts. To account for this, we introduce the formal framework of Many-Valued Logic – also known as ‘Fuzzy Logic in the narrow sense’[1] (Hájek, 2006).

3.1 Introduction to Many-Valued Logic

In contrast to classical logic, Many-valued Logic does not conceive sentences to be either ‘true’ or ‘false’ but allows for a (potentially infinite) number of truth-values in between these values (Priest, 2008). The simplest form of generalizing classical logic in this way is to add a third truth value, which is to be understood as ‘neither true nor false’ (Łukasiewicz, 1920; Post, 1921). The number of truth values can be further extended to an infinite number (Łukasiewicz, 1970), resulting in ‘true’ and ‘false’ as two extremes on a continuum of possible truth-values.

The introduction of Fuzzy-Set Theory by Zadeh (1965) inspired a range of attempts to approach the field of Many-Valued Logic from a set theoretical perspective. So called ‘Fuzzy Inference’ methods found wide applications in various technical fields (mainly Artificial Intelligence and Fuzzy Control), but often lacked a sound mathematical foundation (Hájek, 1998). However, in recent years there have been great advances in the formulation of a coherent mathematical background, which links Many-Valued Logic to Fuzzy-Set Theory based on certain algebraic structures – so called t-norm based algebras (Alsina, 1985; Gottwald and Novák, 2000; Gottwald and Hájek, 2005; Novák, 2012). Drawing on this work, we now have a correspondence between Many-Valued Logic and Fuzzy-Sets quite in the way we find it between two-valued logic and classical set theory. Therefore, in order to enhance comprehensibility, the following section introduces the formalism of Many-Valued Logic using the more intuitive concepts of Fuzzy-Set Theory. To avoid unnecessary formulae, mathematical notation is replaced by (or at least supplemented with) an explanation in plain English whenever possible.

[1] The formalism used within this paper is to be distinguished from ‚Fuzzy Logic in the wide sense‘, which covers a wide variety of methods based on Fuzzy Set Theory as introduced by Zadeh (1965). In contrast to these methods, ‘Fuzzy Logic in the narrow sense’ refers to a branch of mathematical logic. See Hajek (2006) for further elaboration of this distinction.

3.2 Fuzzy-Sets

A Fuzzy-Set is constituted by a class of objects of a certain domain and a function, which assigns a degree of membership to every object in the domain. The degree of membership is usually a real number between 0 and 1, 0 meaning 'no membership' and 1 meaning 'complete membership'. It follows from this definition, that a classical (also denoted as 'crisp') set is a special case of a fuzzy set where every object is assigned either full membership (1) or no membership (0). Formally, we usually denote a fuzzy set with an uppercase letter. The degree of membership of a fuzzy set A is written as m_A. To further simplify the notation it is assumed that the domains of different fuzzy sets A, B, C ... are part of a universal domain X. Thus, the membership degrees of A, B, C ... can each be specified as a function of X.

The sum of all membership degrees of a fuzzy set A is called its *scalar cardinality* (written as $|A|$). A fuzzy set is called *empty*, if its scalar cardinality equals 0 (i.e. all of its membership degrees are 0). Two fuzzy sets A and B are *equal* (written as $A = B$), if the membership degrees of all their elements are equal. A fuzzy set A is called a *subset* of B (written as $A \subseteq B$), if for every object it holds that $m_A \leq m_B$.

The *complement* of a fuzzy set A (written as $\bar{A}$) can be calculated by subtracting the membership of every object from 1. The *intersection* of two fuzzy sets A and B (written as $A \cap B$) can be defined as the minimum of the membership degrees for every object in A and B. And finally, the *union* of two fuzzy sets A and B (written as $A \cup B$) is given by the maximum of the membership degrees for every object in A and B.

Fuzzy set theory as outlined above constitutes a generalization of classical set theory in the sense that all classical set operations (i.e. intersection, union and complement) can be seen as special cases of the corresponding fuzzy operations limited to the membership degrees 0 and 1 (Klir and Yuan, 1995).

3.3 Defining psychological constructs

Within Fuzzy-Set Theory it is possible to define even vague psychological construct in a mathematically explicit manner (cf. Belohlavek and Klir, 2011). Additionally, it provides a comparably simple framework to account for semantic relations between different constructs.

Formally, this is accomplished by identifying a *construct meaning* with a fuzzy-set on a domain of well-defined *attributes.* Thus, a construct can be fully defined by the membership of a set of attributes to the corresponding construct meaning. If these attributes are assessed by the items of a psychological test, the meaning of a construct corresponds to the degree of membership each test item bears to the construct.

If several constructs are defined on the same domain of attributes, their semantic relations can be described by means of fuzzy set operations. The complement of a construct meaning semantically denotes the *opposite* of that construct. The union of two construct meanings can be interpreted as the composite meaning of both construct. The scalar cardinality of a construct meaning denotes its *semantic content*. Fuzzy equality resembles *synonymy*, and the intersection of two construct meanings resembles the semantic content that is shared by both constructs, i.e. the shared meaning of two constructs. The scalar cardinality of this shared meaning is called the *semantic overlap* between two (or more) constructs. Moreover, the subset relation $\subseteq$ can be interpreted as a *sub-concept* relation, constituting a partial ordering of construct meanings. The following example shall illustrate these ideas.

Example 1: *A domain of well-defined attributes be specified by the following test items:*

- *a) I like to meet new people.*
- *b) I like to go to parties.*
- *c) I am used to being the center of attention.*
- *d) I have many friends.*
- *e) Most people I know think I am a likable person.*

Given this domain of attributes it is possible to define a range of constructs by assigning corresponding membership degrees to these attributes .Table 1 shows the membership degrees of a collection of construct meanings defined on the specified domain.

Table 1: Example of construct meanings defined on a common domain of Attributes

Construct	**a**	**b**	**c**	**d**	**e**
Social Orientation	0.9	0.7	0.6	0.8	0.8
Open Mindedness	0.6	0.2	0	0.2	0.2
Popularity	0	0	0.5	0.7	0.8
Extraversion	0.8	0.7	0.4	0.2	0

The semantic relations between these construct meanings are illustrated by an Euler diagram in Figure 1. The ellipses represent the semantic content of the corresponding constructs. The areas of the ellipses are proportional to their semantic content. The areas of the overlap between adjacent ellipses are proportional to the semantic overlap of the corresponding constructs. It can be seen that (given the formal definition used in this example) open mindedness*,* popularity *and* extraversion *are all sub-concepts of a more general construct called* social orientation*. Moreover, the diagram shows the semantic overlap between these three constructs. For example, it becomes clear that* popularity *is semantically closer to* extraversion *than to* open mindedness*. Additionally,* open mindedness *and* extraversion *share a considerable amount of semantic content, whereas the intersection of all three concepts is rather small.*

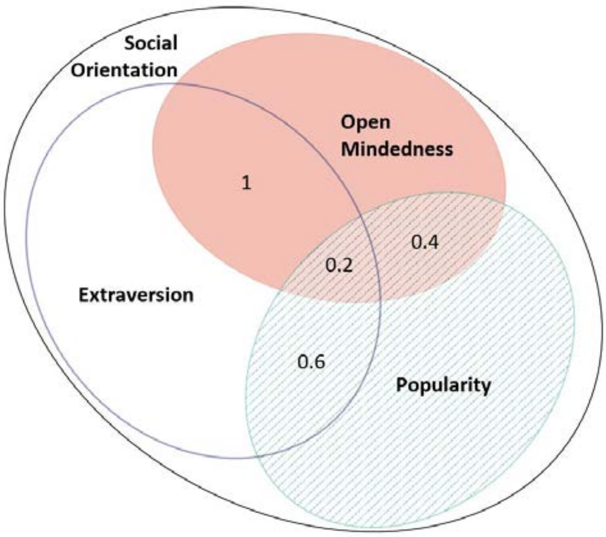

Figure 1 Euler Diagram showing the semantic relations of the construct meanings defined in Table 1.The areas of the ellipses are proportional to the corresponding semantic content (i.e. the scalar cardinality of the underlying fuzzy set). The numbers denote the degree of semantic overlap between two adjacent constructs (i.e. the scalar cardinality of the intersection of the underlying fuzzy sets). The graph was generated using the eulerAPE software by Micallef and Rodgers (2014)

4 Reconstructing the Logic of psychological Tests

The (fuzzy-) set theoretical formulation of a construct meaning serves as a semantic basis to give necessary and sufficient conditions for the inference of a construct from the answers in a psychological test. Using the formalism of many-valued logic it is now possible to reconstruct psychological testing procedures as a standardized way of inference. The next sections start with a brief introduction to a special type of many-valued logic (Quantified Łukasiewicz Logic), followed by a re-interpretation of psychological testing as logical inference and its theoretical implications.

4.1 Quantified Łukasiewicz Logic

The following calculations are based on a many-valued extension of classical predicate calculus called Quantified Łukasiewicz Logic ($Ł_\forall$)[2]. The calculus goes back to the work of Łukasiewicz (1970) and was adapted and further developed by Hájek (1998b). Because there is a close resemblance between the syntax of classical predicate logic and $Ł_\forall$, we focus on the semantic aspects of the calculus.

The logical connectives used in $Ł_\forall$ are → (implication), ≡ (equivalence), ∧ (logical AND), ∨ (logical OR), & (strong logical AND) and ¬ (negation)[3]. To quantify expressions the calculus uses the signs ∀ (for all) and ∃ (there is at least one). Furthermore, there are *predicates* (P, Q, R, …) which are used to express relations between *object constants* (c, d, …) or *object variables* (x, y, …). Object constants and object variables are also referred to as *terms*. These terms may be of different *sorts* (s_1, s_2 …). Each predicate is said to be of a certain *type* depending on a corresponding sequence of sorts ($s_1, \dots, s_n$) for which it can be used. Additionally, $Ł_\forall$ makes use of two truth constants: $\overline{1}$ and $\overline{0}$, to syntactically account for truth and falsity.

Expressions are defined recursively: *atomic expressions* consist of a predicate and a number of terms (either logical constants or logical variables) and take the form $P(t_1, t_2, \dots)$. If φ, ψ are expressions and x is an object variable, then $\varphi \to \psi, \varphi \wedge \psi, \varphi \vee \psi, \varphi \& \psi, \varphi \equiv \psi, \neg\varphi, (\forall x)\varphi, (\exists x)\varphi, \overline{0}, \overline{1}, \dots$ are expressions, too. The rule of inference used in $Ł_\forall$ is the modus ponens (Hajek, 1998b).

Semantically, any predicate in $Ł_\forall$ can be identified with a fuzzy relation (i.e. a fuzzy subset of the cross product of the corresponding terms). The truth value of a logical expression φ is designated as $\text{tv}[\varphi]$ The semantics of $Ł_\forall$ further consist in assigning truth functions to the basic logical operations (these are generalizations of the truth tables used in classical logic). The truth functions used in $Ł_\forall$ resemble the corresponding (fuzzy-) set theoretical operations introduced in section 3.2. Negation corresponds to the fuzzy complement, logical AND to fuzzy intersection, logical OR to fuzzy union and equivalence corresponds to fuzzy equality. We thus arrive at the following truth functions:

$$f_\neg(x) = 1 - x \tag{4.1}$$

$$f_\wedge(x, y) = \min(x, y) \tag{4.2}$$

$$f_\vee(x, y) = \max(x, y). \tag{4.3}$$

Additionally, the semantics of ∀ and ∃ are given by the least upper bound (inf) and greatest lower bound (sup) of the quantified logical variable, respectively. The truth function of & is given by the so called Łukasiewicz t-norm:

$$f_\&(x, y) = \max(0, x + y - 1). \tag{4.4}$$

Finally, logical implication is semantically represented by

$$f_\to(x, y) = \min(1, 1 - x + y). \tag{4.5}$$

This function is also known as the *residuum* of the corresponding t-norm. Therefore, the semantics of $Ł_\forall$ can also be defined in the form an algebraic structure known as *residuated lattice* (see Hájek, 1998b for details).

Following Hájek (1998b) it is possible to further enhance the expressive power of the calculus by adding the concept of so called *generalized quantifiers*, which syntactically act in much the same way as ∀ and ∃. The core idea is to generalize the semantic representation of a quantifier to a function of

---

[2] Actually, the calculus introduced is a *many-sorted* variant of the actual Łukasiewicz logic with *generalized quantifiers*.

[3] In fact, the basic alphabet of $Ł_\forall$ consists of fewer signs from which the remaining ones are defined (Priest, 2008). However, the distinction between basic connectives and derived connectives is not necessary for the purposes of this paper and is therefore omitted.

the truth values of the corresponding basic elements. For ∀ and ∃ these functions are inf and sup, respectively, whereas for different quantifiers there may be other functions.

A quantifier which will be of particular interest in the following sections is the quantifier $Many$, which is semantically represented by relative frequencies (in the crisp case) or – more generally – by the arithmetic mean of the corresponding truth values[4].

Using these truth functions it is possible to calculate the truth value of any valid expression in $Ł_\forall$ from its corresponding atomic expressions, resulting in a both *sound* (i.e. every provable expression is semantically true) and *complete* (i.e. every semantically true expression is provable) calculus (Hájek, 1998b).

4.2 Inferring psychological Constructs using Many-Valued Logic

The following sections use a formal language based on $Ł_\forall$ to express some basic assumptions about the inherent logic of psychological testing procedures. These assumptions rely on the idea that constructs are vague concepts in the sense given in section 3.3. The resulting formal expressions are then used to derive the truth values of certain expressions relevant to the interpretation of psychological tests applying the corresponding truth functions given in section 4.1.

4.2.1 The Logic of psychological Tests

The first step towards a logical account of psychological tests is to introduce some basic definitions. We start with the introduction of four different sorts of objects in our domain of discourse: Attributes, constructs, tests and persons. These sorts are denoted A, C, T and P, respectively. If a term t is of a certain sort s_i, this is written as t_{s_i}. Thus, if, for example, c is an Attribute, we express this by writing c_A.

We further define the following predicates:

- $D(t_A, t_C)$: 'defines' (relating attributes to constructs)
- $I(t_A, t_T)$: 'in' (relating attributes to tests)
- $P(t_P, t_A)$: 'possesses' (relating persons to attributes)
- $H(t_P, t_C)$: 'has' (relating persons to constructs)

With this basic vocabulary it is now possible to define necessary and sufficient conditions for correctly ascribing a construct to a person:

$$\forall x_A: \big(D(x_A, i_C) \rightarrow P(j_P, x_A)\big) \equiv H(j_P, i_C). \tag{4.6}$$

In plain text this means: A person j is said to have construct i if and only if she possesses all attributes x which define the construct i.

Applying this definition, it follows that – in order to find out to which degree a person has a construct – a psychological test needs to assess to which degree the defining attributes of the construct are met. This, of course, is only possible, if a test contains these attributes. To account for this, we define a test k to be *strictly valid* for a construct i (written as $V(k_T, i_C)$), if and only if it contains the defining attributes of i:

$$V_{strict}(k_T, i_C) \equiv \forall x_A: \big(D(x_A, i_C) \rightarrow I(x_A, k_T)\big). \tag{4.7}$$

We further say a person j *strictly satisfies* the conditions for ascribing construct c assessed by a test k (written as $S(j_P, i_C, k_T)$) if she possesses all attributes x which both define construct i and are contained in test k:

$$S_{strict}(j_P, i_C, k_T) \equiv \forall x_A: ((D(x_A, i_C) \wedge I(x_A, k_T)) \rightarrow P(j_P, x_A)). \tag{4.8}$$

[4] In fact, the definition of $Many$ is not quite as straightforward as presented here. But since this paper is merely concerned with the application of logical calculus rather than mathematical developements, we spare the technical details here (see Hájek, 1998b for a complete presentation).

One might argue, however, that these conditions are too restrictive to be applied to real world settings. Do we really demand *each and every* defining attribute to be possessed by a person to correctly ascribe a certain construct? And do we only speak of a valid test, if it contains *all* defining attributes? In many cases, we would resent from this strong formulation. Therefore we introduce a weaker version of the above formulae, using the quantifier $Many$ instead of ∀. We thus define *test validity* as:

$$V(k_T, i_C) \equiv Many\ x_A: \big(D(x_A, i_C) \rightarrow I(x_A, k_T)\big) \tag{4.9}$$

and approximate satisfaction of the test conditions accordingly by:

$$S(j_P, i_C, k_T) \equiv Many\ x_A: ((D(x_A, i_C) \land I(x_A, k_T)) \rightarrow P(j_P, x_A)). \tag{4.10}$$

In this weaker formulation we do not expect a person to possess all attributes in order to ascribe the corresponding construct. Instead it suffices, if the person possesses many of the defining attributes. The same holds for the definition of approximate test validity.

From these definitions it follows immediately, that if a test k is valid for a construct i and a person j satisfies the conditions of test k, she has the construct under consideration:

$$(V(k_T, i_C) \land S(j_P, i_C, k_T)) \rightarrow H(j_P, i_C). \tag{4.11}$$

In everyday language this expression can be paraphrased as follows: *If a test is valid for the assessment of a certain construct and a person possesses many of the attributes contained in the test, one can infer that she has the construct under consideration.*

4.2.2 Calculation of Construct Scores

To give a semantic interpretation of the above formulae we identify $D(t_A, t_C)$ with a *construct definition* in the sense of section 3.3. The predicate $P(t_P, t_A)$ is interpreted as the empirical *testing behavior* of a person (conceived as the degree to which test items apply to a person[5]). Finally, we take $I(t_A, t_T)$ to be a crisp relation indicating whether an attribute is part of a certain test or not.

Since the truth value of equation 4.11 is 1 by definition, it follows from the truth function of → that

$$\mathrm{tv}[V(k_T, i_C) \land S(j_P, i_C, k_T)] \leq \mathrm{tv}[H(j_P, i_C)]. \tag{4.12}$$

Given a valid test we can thus infer a *lower bound* of the construct score $\mathrm{tv}[H(j_P, i_C)]$ from the test behaviour of a person.

Applying the truth function of ∧ to this inequality, we get

$$\min(\mathrm{tv}[V(k_T, i_C)], \mathrm{tv}[S(j_P, i_C, k_T)]) \leq \mathrm{tv}[H(j_P, i_C)]. \tag{4.13}$$

This implies that the lower bound of $\mathrm{tv}[H(j_P, i_C)]$ is both limited by the degree to which a person possesses the attributes assessed in the test (i.e. $\mathrm{tv}[S(j_P, i_C, k_T)]$) on the one hand and the test's *validity* on the other hand.

By means of the corresponding truth functions, equation 4.9 yields:

$$\mathrm{tv}[V(k_T, i_C)] = \frac{1}{c_A} \sum \min(1, 1 - \mathrm{tv}[D(x_A, i_C)] + \mathrm{tv}[I(x_A, k_T)]), \tag{4.14}$$

[5] The test behaviour is assumed to be a number between 0 and 1 – to achieve this, usually a linear transformation of test answers will be applied. However, other transformations may be justifiable depending on the context.

with c_A being the cardinality of the domain set A[6].

Note that, since $\text{tv}[I(x_A, k_T)]$ is a crisp relation, it follows that

$$\min(1,1 - \text{tv}[D(x_A, i_C)] + \text{tv}[I(x_A, k_T)]) = \begin{cases} 1, & \text{if } x \text{ is in } k \\ 1 - \text{tv}[D(x_A, i_C)], & \text{if } x \text{ is not in } k. \end{cases} \quad (4.15)$$

Since for all attributes contained in the test, the leftmost side of equation 4.15 equals 1, one can use the number of test items to facilitate the calculation. Adding the sum of the remaining residua to this number and dividing by c_A one easily obtains a test's valitidy.

To obtain the value of $S(j_P, k_T)$, we apply the corresponding truth functions to equation 4.10:

$$\text{tv}[S(j_P, i_C, k_T)] = \frac{1}{c_A} \sum (\min(1,1 - \min(\text{tv}[D(x_A, i_C)], \text{tv}[I(x_A, k_T)])) + \text{tv}[P(j_P, x_A)]), \quad (4.16)$$

with c_A being the cardinality of the domain set A.

Again, we can simplify this calculation due to $I(x_A, k_T)$ being a crisp relation: If an attribute is not contained in the test, $\min(\text{tv}[D(x_A, i_C)], \text{tv}[I(x_A, k_T)])$ becomes 0, resulting in

$$\begin{aligned} &\min(1,1 - \min(\text{tv}[D(x_A, i_C)], \text{tv}[I(x_A, k_T)])) + \text{tv}[P(j_P, x_A)]) \\ &\quad = \min(1,1 + \text{tv}[P(j_P, x_A)]) = 1. \end{aligned} \quad (4.17)$$

The calculation therefore reduces to adding the number of attributes not contained in the test to the residual sum from the attributes which are in the test and dividing the result by the number of relevant items[7].

Inserting this in equation 4.13 we finally obtain:

$$\min \begin{pmatrix} \frac{1}{c_A} \sum \min(1,1 - \text{tv}[D(x_A, i_C)] + \text{tv}[I(x_A, k_T)]), \\ \frac{1}{c_A} \sum (\min(1,1 - \min(\text{tv}[D(x_A, i_C)], \text{tv}[I(x_A, k_T)])) + \text{tv}[P(j_P, x_A)]) \end{pmatrix} \leq \text{tv}[H(j_P, i_C)]. \quad (4.18)$$

Applying the semantics of $Ł_\forall$ to the logical reconstruction formulated in section 4.2.1 it is now possible to calculate the lower bound of $\text{tv}[H(j_P, i_C)]$ from the formal definition of a construct and the test behavior of a person. This value semantically corresponds to the *degree to which a construct can be inferred from a person's test behavior* and shall therefore be referred to as the *construct score* of a person.

Example 2: *Based on the formal definition of 'Extraversion', we imagine a test consisting of only three items:*

a) I like to meet new people.
b) I like to go to parties.
c) I am used to being the center of attention.

Note that the items d) and e) have been removed from the test, because their membership degrees for the construct 'Extraversion' are rather small. Table 2 shows the hypothetical test results of a person (Peter) alongside with the corresponding membership degrees.

[6] Usually, we will only include attributes in which are related to the construct at hand (i.e. have a membership degree greater than zero). However, from a formal point of view, this is not necessarily the case.

[7] The fact that attributes which are not assessed in the test always evaluate to 1 (irrelevant of the answer a person may have given) may seem counterintuitive: the less attributes the test assesses, the more a person satisfies the tested conditions. However, since every attribute not contained in the test reduces its validity, this does not impose any practical limitations.

Table 2: A hypothetical test result for the construct 'Extraversion'. D refers to the construct definition, P to the degree to which the tested person possesses the corresponding attribute.

Item	***D***	***P***
I like to meet new people.	0.8	0.7
I like to go to parties.	0.7	0.5
I am used to being the center of attention.	0.4	0.7

To calculate Peter's construct score we insert the corresponding values of D and P into the equations 4.14 *and* 4.16. *Note that there is one additional attribute related to the construct, which is not included in the test: Item d) with a membership degree of 0.2, hence c_A equals 4.*

$$\begin{aligned}\text{tv}[V(k_T, i_C)] &= \frac{1}{c_A}\sum \min(1, 1 - \text{tv}[D(x_A, i_C)] + \text{tv}[I(x_A, k_T)]) \\ &= \frac{1}{4}(\min(1, 1 - 0.8 + 1) \\ &+ \min(1, 1 - 0.7 + 1) \\ &+ \min(1, 1 - 0.4 + 1) + \min(1, 1 - 0.2 + 0)) \\ &= \frac{1}{4}(\min(1, 1.3) + \min(1, 1.2) + \min(1, 1.6) + \min(1, 0.8)) \\ &= \frac{3.8}{4} = 0.95\end{aligned} \tag{4.19}$$

and

$$\begin{aligned}\text{tv}[S(j_P, i_C, k_T)] &= \frac{1}{c_A}\sum(\min(1, 1 - \min(\text{tv}[D(x_A, i_C)], \text{tv}[I(x_A, k_T)])) + \text{tv}[P(j_P, x_A)]) \\ &= \frac{1}{4}(\min(1, 1 - \min(0.8, 1) + 0.7) \\ &+ \min(1, 1 - \min(0.7, 1) + 0.5) \\ &+ \min(1, 1 - \min(0.4, 1) + 0.7) + \min(1, 1 - \min(0.2, 0) + 0)) \\ &= \frac{1}{4}(\min(1, 1 - 0.8 + 0.7) \\ &+ \min(1, 1 - 0.7 + 0.5) + \min(1, 1 - 0.4 + 0.7) + \min(1, 1)) \\ &= \frac{1}{4}(0.9 + 0.8 + 1 + 1) = \frac{3.7}{4} = 0.93.\end{aligned} \tag{4.20}$$

Since the degree to which Peter satisfies the assessed conditions for calling him extraverted is smaller than the test's validity, inserting the results into equation 4.18 *finally yields*

$$0.93 \leq \text{tv}[H(j_P, i_C)].$$

We thus obtain a construct score of 0.93, which means that the test result enables us to infer Peter to be an extraverted person to the degree 0.93 – or, to put it in other words: The degree to which Peter can be correctly ascribed the construct 'Extraversion' is at least 0.93.

5 Main Results

Starting with a formal representation of construct meanings within the framework of fuzzy set theory, the above sections have presented a rational reconstruction of psychological testing procedures as a standardized way to infer a person's construct score from her testing behavior. The use of formal logical calculus enabled the algebraic calculation of construct scores on the basis of many-valued semantics. The main result of this reconstruction is a scoring algorithm for psychological tests, which is consistent with the interpretation of psychological constructs as abstract concepts (rather than latent variables).

The scoring algorithm derived in section 4.2.2 can be paraphrased as follows:

1. Compare the answer for each test item with the degree to which the corresponding attribute belongs to the construct under consideration: If a person scores at least as high as the corresponding membership degree, the person gets an *item score* of 1 for this item; if the person scores lower than the corresponding membership degree, the item score is calculated by subtraction of the difference between test behavior and membership degree from 1. For relevant items not contained in the test, the corresponding item score is also 1.
2. For a strictly valid test (i.e. a test which contains all defining attributes of the construct) the mean item score marks the *lower bound* of the person's construct score.
3. If *not* all of the defining attributes are contained in the test, calculate the test's validity: For each attribute not assessed in the test, take the membership degree and subtract it from 1. Sum up the resulting values and add the number of attributes the test does assess. Finally, divide the result by the number of all relevant attributes. If the result is smaller than the mean item score, it marks the lower bound of the construct score.

The resulting construct score semantically refers to the degree to which a construct can be inferred from a person's test behavior given the test's validity and the corresponding construct definition. This value can be interpreted as the lower bound of the degree, to which a construct is suitable to describe the tested person. A high construct score means that it is semantically correct to ascribe the construct to the tested person, whereas a low construct score means that the test gives no evidence for the ascription of the construct to that person.

In accordance with the concept view of psychological constructs, construct scores obtained from a test do not refer to latent variables but to the truth value of the expression 'person j has construct i'. Therefore, test results are not interpreted as measurements but as the result of formal logical inference.

6 Practical Implications

Because the approach presented in this paper does not make any assumptions about the empirical structure of latent variables, psychological traits or underlying unobservable phenomena, it offers a wide range of possible applications. Especially when there is little empirical or theoretical evidence on the empirical structure of a latent variable, the semantic approach outlined above offers a reasonable alternative to conventional test models. In some cases, there may even be doubts about the existence of a psychological trait. Moreover, in many diagnostic contexts, the latent structure of testing behavior bears no relevance at all. This results in a wide range of possible applications. Therefore, it is worthwhile to investigate how the developed approach can be implemented to practically guide test construction and test revision.

6.1 Test construction

Conventional psychometric methods of test construction usually rely on statistical aspects of a test, such as the covariance structure of test items or correlations between different tests (cf. Kline, 2000). In contrast to these procedures, the approach presented here builds entirely on a formal definition of the construct under consideration. Thus, the *meaning* of a construct becomes the focus of attention.

In order to specify a construct meaning in the above sense, there are two basic rationales that can be followed. Firstly, one can try to analytically decompose an existing construct meaning into its defining attributes. And secondly, one can start with a domain of attributes (or test items) and search for empirically meaningful constructs using a synthetic approach.

6.1.1 Analytical Approach

The main task within the analytic approach is to identify a set of attributes as the semantic basis of a construct and to assign membership degrees to these attributes. The easiest way to accomplish this is to assess expert ratings for the semantic relations between each attribute and the construct to be defined. These expert ratings can be either direct (i.e. numbers between 0 and 1), or the result of an indirect scaling procedure (e.g. pair comparison procedures)

An expert query is relatively easy to conduct and yields reasonable results in most cases. However, its results may vary depending on who is considered to be an 'expert'. The resulting definitions usually reflect the opinion of a very limited number of people. Therefore, expert ratings are especially suitable if one is concerned with the assessment of Traits based on a number of specific criteria, such as in contexts of employee selection.

Instead of querying experts, one might just as well assess membership degrees by means of semantic networks, which are based on word associations or actual language use of a whole population. The distances in such a semantic network may serve as membership degrees in order to define the construct at hand. This approach applies well to constructs which are based of everyday language, such as personality traits (compare Buss and Craik, 1980, 1981; Buss, 1983) or emptions (see Russell and Fehr, 1994).

Analytic methods of test constructions are, in a way, an attempt to *analyze actual language use*. Thus, their results are always to be interpreted against the background of a certain speech community.

6.1.2 Synthetic Approach

A complementary approach to define a construct is to take a set of attributes and construct suitable membership degrees which yield empirically meaningful interpretations. The basic rationale behind this is to find a small set of distinctive constructs which are able to differentiate well between the members of a given population. There are various algorithms which can be exploited for this purpose, like for example a search for most frequent item sets (compare Tan, Steinbach and Kumar, 2006 for an overview).

If, however, one rather aims at a relative comparison between the members of a given population, it may be more reasonable to construct membership degrees from relative frequencies of item answers. For example, it may be reasonable to take the most frequent item answer as a membership degree for the corresponding attribute, resulting in a construct which reflects the combination of the most typical item answers. Similarly it is possible to define a 'median construct' which takes the median answer of each item as a membership degree (or, more generally, 'percentile constructs'). This would resemble the totality of maximum (or minimum) item answers which are given by a predefined percentage of the population of interest.

Synthetic test construction is best to be understood as a way of *concept building*. Therefore, any method applied must ultimately be evaluated by means of theoretical fruitfulness and empirical utility.

6.1.3 Item Selection

In many cases it is desirable to reduce test length to a reasonable amount of items. To accomplish this it is possible to exclude items which give relatively little information about the construct of interest.

Item selection should aim at an optimal compromise between test length and test *validity*.

Since removal of attributes from the test both results in higher mean item scores as well as lower validity values, this is not a trivial task. We suggest an empirical approach, based on a representative sample from the relevant population. Using an iterative procedure starting with the removal of the item with the lowest membership degree (and continuing with the next lowest and so forth) it is possible to optimize for the highest mean construct score of the sample at hand.

This procedure is quite similar to standard item selection based on optimization of Cronbach's $\propto$. However, the optimization criterion is completely different. Instead of minimizing the negative effects of presumed measurement error the above algorithm tries to maximize the degree to which constructs can be inferred in a given population.

6.2 Interpretation of Test Scores

It should be stressed that, despite taking a numerical value between 0 and 1, the construct scores obtained by the above calculations are *not* probabilities. The superficial similarity tempts to misinterpret test scores in a Bayesian fashion (i.e. the probability, that a person has a certain trait). To confound test scores with probabilities is wrong both in form and content. Since the semantics of many-valued logic do not conform to the Kolmogorov axioms, they are definitely no probabilities. Additionally, in contrast to probabilities, truth-values do not express uncertainty about the world. Instead, they describe *vague semantic relations* – even if everything was known about the attributes a person possesses, there would still be constructs that do not fully apply to her, therefore giving test scores below 1.

Another temptation lies in conceiving test scores as measures of psychological traits. However, in the approach at hand, a test score is a *logical truth-value*, reflecting the degree to which a construct yields an adequate description of a person. This truth-value is inferred from the subjective ratings obtained in a test using a formal definition of the construct to be assessed. Since the formalism presented makes no assumptions about an underlying trait structure, test scores are to be interpreted in purely semantic terms. Thus, it does not tell us anything about psychological variables or cognitive processes involved in test behavior. Instead, a test score tells us, to what extent a person fulfills the necessary and sufficient criteria for the *ascription* of a certain construct.

7 Limitations

7.1 Uniqueness of Test Scores

Since there are various types of Many-Valued Logic, the question of uniqueness naturally arises. In fact, there is a potentially infinite set of different semantics that could have been applied in the above sections (cf. Gottwald, 2010). However, most of the corresponding logical calculi can be conceived as special cases of the same axiomatic basis for logics based on continuous t-norms, the so called Basic Fuzzy Logic (BL) (Hájek, 1998a). An important fact is that the truth functions of $\wedge$ and $\vee$ are always given by the minimum and maximum operators, regardless of the semantics implied by a special case of BL. Moreover, any t-norm based implication operator yields one if and only if the truth value of the antecedent is smaller or equal to the truth value of the consequence and satisfies the monotonicity condition (Alsina, 1985). Thus, as long as construct scores are calculated based on the $\forall$-quantifier, the results are unique up to monotone transformations with regard to the family of $BL_{\forall}$-logics. If, however, one choses the less restrictive quantifier $Many$ it should be noted that further assumptions regarding the chosen semantics are implied.

7.2 Subjective Truth Values

One could argue that although there is no measurement theory involved in the above formalism, the application to actual test behavior does presume item answers to be assessed on an interval scale level. Although this objection is true from a formal point of view, it does not follow that the above formalism implicitly models a measurement relation. In order to be comparable to a construct definition, test answers have to be real numbers between 0 and 1, reflecting the subjective truth-values of the corresponding attributes for the tested person. This assumption does indeed make assumptions about an underlying measurement structure. However, these only refer to the item level and do not extend to theories about latent variables. Therefore, we are concerned with a *scaling* problem. In fact, if we measure anything at all, it is always the same 'psychological' variable: the agreement to certain test items. It might be possible to apply (possibly probabilistic) scaling models in order to justify these assumptions. However, the correct use of a metric scale in a test is a far weaker claim than to measure latent traits expressed by test behaviour. Hence, the semantic approach presented above gives a more parsimonious account of psychological testing than traditional psychometric theories.

7.3 Definability of Constructs

Finally, one may object to the claim, that psychological constructs can be decomposed into 'well defined attributes' at all. If one conceives natural languages as intrinsically vague, this is indeed a valid objection. From a practical point of view, this results in the problem of item construction. It is never easy to find unambiguous items. In many cases, the ideal of a set of well-defined attributes will hardly be reached. However, even if the attributes used to define a construct are vague themselves, this does not mean that it is useless to attempt a semantic decomposition of psychological constructs. In any case test items will normally be less vague than the constructs under consideration, thereby reducing fuzziness and semantic ambiguity, even if they are not truly 'well-defined'.

In theory one could also attempt to model the vagueness of items in a similar way as we modeled vague constructs. However, this could either lead to an infinite regress (since the results will probably be vague as well), or lead to the construction of test items which are more precise than the knowledge of the tested person. Moreover, in some cases the number of items needed for absolute precision would be potentially infinite. If, for example, the item 'I have many friends' was decomposed into one item for each number of friends, the result would be an infinite set of items.

Another important point is, whether we actually want perfect precision (i.e. all tested people have the exact same understanding of the items) in a psychological test. In many cases the more relevant information may be to assess certain *self-ascriptions*. The question at hand is, whether psychological constructs are ascribed due to objectively observable facts or at least in part based on personal constructions of the tested person.

In conclusion, one may say that even if psychological constructs are not completely decomposable into well-defined attributes, their meaning may be clarified by the reduction of vagueness through a set of attributes. Moreover, in some cases it may even be desirable to keep a certain degree of vagueness when it comes to the formulation of test items.

Finally, it should be noted that this is not a problem unique to the approach at hand, but a general problem of test construction. In fact, one could even argue that one of the strengths of the current approach is that it strips psychometric theory down to a point where these general problems of test constructions (like item formulation or scale use) become so obvious that they can no longer be ignored or covered by the use of sophisticated statistics.

7.4 Qualitative vs. quantitative Constructs

The formalization presented above conceptualizes psychological constructs as *qualitative* concepts. The degree to which a construct is ascribed to a person may vary in degree, depending on the derived truth value. However, the test score does not refer to the quantitative value of an underlying trait. Constructs in the above sense are taken to be qualitative linguistic concepts which apply to different people to varying degrees.

Naturally, the question arises what to make of quantitative constructs – do they lie outside the scope of the presented formalism, are they more complex linguistic constructions, or maybe they do not exist at all? We shall respond to these questions in the following paragraph.

One way of extending our approach to quantitative variables is to interpret psychological traits as *linguistic variables* (Zadeh, 1975a, 1975b, 1975c). A linguistic variable consists of a quantitative variable (e.g. 'Extraversion'), a set of linguistic terms (e.g. 'slightly extraverted', 'rather extraverted' etc.) and a function for each term which maps the quantitative variable to the corresponding fuzzy membership degrees (compare Fig. 2). For each linguistic term we can further define a fuzzy set of attributes resulting in a formally defined construct in the above sense. In the case of physical variables like for example 'temperature' the underlying variable and the corresponding membership functions are well-defined. However, this is not the case for latent variables like 'Extraversion'. We could, of course, define the membership function on an arbitrary metric chosen on pragmatic grounds – for example a scale from zero to five like in figure 2. Given such a linguistic variable, it would be possible to calculate the quantitative base value from the scores of the linguistic-term-constructs inferred from a test using the formalism outlined in this paper (this procedure known as 'de-fuzzification' of a linguistic variable).

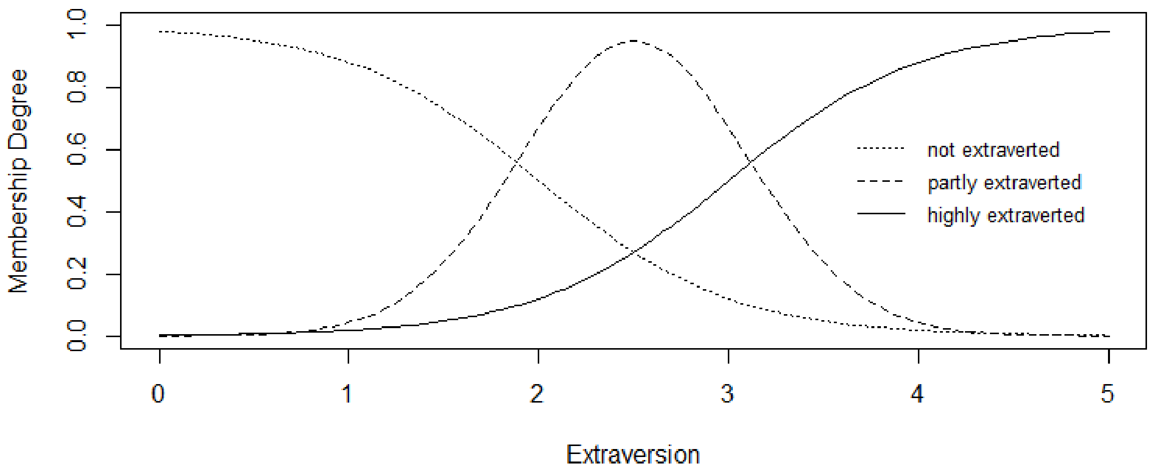

Figure 2: Membership functions of a linguistic variable consisting of a (latent) quantitative variable ('Extraversion') and three linguistic terms ('not extraverted', 'partly extraverted', 'highly extraverted')

However appealing this approach may be, we are very skeptical about its usefulness. In the above case, for example, the Extraversion value might as well be calculated from the construct 'highly extraverted' alone. The 'quantitative' information of the additional constructs is therefore redundant. Moreover, since the absolute numerical value of the scale is arbitrary, we gain no additional information over the use of just one prototypical construct (e.g. 'highest possible amount of Extraversion'). Thus, instead of specifying a set of functions mapping a quantitative variable onto the membership degrees of a set of linguistic constructs, in most cases one may just as well stick to the (more parsimonious) qualitative account specified above.

It is quite an interesting fact that the qualitative/quantitative distinction seems to somehow dissolve when applying the approach presented in this paper. However, we think this is hardly surprising, because the construct view of psychological constructs interprets psychological variables as *linguistic constructions*. Constructions per se bear no necessary resemblance to real entities. Therefore, whether we use qualitative or quantitative constructions merely specifies the way we *talk* about psychological constructs. We should, therefore not expect any further insights in their 'true' nature.

7.5 Relation to other Psychometric Approaches

Although the approach presented in this paper is formulated on a completely different theoretical background, there are some interesting connections to traditional psychometric methods.

A common way to construct psychological tests consists in applying an Exploratory Factor Analysis (EFA) to a set of items. The underlying model describes item responses as a linear combination of a small number of unobserved factors (adding a random error term to account for empirical deviations from the model). In many cases, the resulting loading matrix is used to divide items into homogenous groups, which are thereafter interpreted based on the semantic contents of the grouped items. This last step can be understood as a way of using semantic information from the items to (informally) define the construct at hand. However, opposite to our approach, in EFA the semantic interpretation of test scores does not resemble the underlying model. Furthermore, the EFA model entirely relies on empirical correlations between item answers in a given population. Using this information to infer semantic relations is questionable, because correlations confound semantic and empirical relations. Of course, the correlational matrix of an item set may be used to *define* abstract concepts based on empirical relations (similar to the synthetic approach to test construction sketched above), but then, again the question arises in which way the EFA model resembles or justifies such a definition. We think that, from a formal point of view, it does not.

A completely different approach to latent variable modelling is offered by Item Response Theory (IRT). The aim of IRT models lies in finding a certain empirical structure, which conforms to a certain model equation (Embretson & Reise, 2000). In contrast to our approach IRT does not include semantic information at all. The empirical structures modeled by IRT try to resemble measurement theoretic properties of (mostly) quantitative variables. The semantic approach presented above, however, does not impose any restrictions on the empirical structure of item responses. Therefore, IRT models do not interfere with our approach – neither on theoretical nor on formal grounds. As mentioned before, in our view it is possible to use a formally defined language provided by psychological constructs without any assumptions about latent variables. As long as one aims at describing people, this is by far the easiest and most reasonable way to go. If, on the other hand, one tries to justify quantitative judgements based on a measurement theoretical empirical structure, IRT models may be more appropriate.

Finally, we would like to address some issues about the so called Classical Test Theory (CTT) in the context of the concept view of psychological constructs. CTT can be understood as a theory of measurement error, independent of actual empirical structures (cf. Gulliksen, 1950; Lord & Novick, 1968). As mentioned in section 2.2 the concept of measurement error is not to be confounded with our notion of semantic vagueness. Additionally, we doubt that it is reasonable to use it as an account of response variation in psychological tests. Since CTT does not offer a satisfying treatment of measurement, either (cf. Embretson & Reise, 2000, we see little justification for its use as a psychometric tool.

Although traditional psychometrics differs from our approach in both theory and formalization, they are closely related on a purely pragmatic level. Looking at equation 4.15 it can easily be seen that if all defining attributes are assessed in the test, the test's validity is one. Additionally, it follows from equation 4.16 that, if all defining attributes fully belong to the construct (i.e. all membership degrees are 1) and the test's validity is 1, the degree a person satisfies the conditions for ascribing the construct at hand reduces to the arithmetic mean of the item answers:

$$\frac{1}{c_A}\sum \mathrm{tv}[P(j_P, x_A)] \leq \mathrm{tv}[H(j_P, i_C)]. \tag{7.1}$$

Given this result it follows that whenever a test score is calculated as the arithmetic mean of item responses, it can be interpreted as a special case of the semantic approach presented in this paper, where all defining attributes fully belong to the construct and no attributes are missing in the test.

Interestingly, mean item responses are the most common statistic used to calculate test results. This applies both to CTT and EFA approaches and even most IRT applications, since in Rasch-Models a person's ability may be expressed by her sum score, which is merely a linear transformation of the scale mean. Therefore, most existing tests can be re-interpreted in the sense of our semantic approach. This even applies to hierarchical constructs like intelligence test batteries which calculate a general IQ from the mean of the contained subscales.

Thus, the logical formalization of psychological testing given in this paper offers an alternative way to make sense of existing tests, bridging the theoretical position of constructs being linguistic concepts with the common practice of calculating mean test scores. At the same time the formal explication of such concepts as 'content validity' or 'construct meaning' results in a more flexible (and possibly more appropriate) way to derive construct scores.

8 Conclusions

This paper introduced a novel approach to psychological tests using the formalism of Many-Valued Logic. Based on the interpretation of psychological constructs as abstract concepts, it was shown that psychological testing procedures can be coherently interpreted as a way of formal inference. Applying formal logical calculus to model the inference from testing behaviour to a construct, a new scoring algorithm for psychological tests was derived. The resulting construct scores reflect the degree to which a person fulfills the defining criteria of a construct, avoiding speculations about underlying latent variables.

In contrast to common psychological procedures, the approach presented here focuses on the *semantic content* of constructs. Whereas common statistical procedures are not concerned with the meaning of test items at all, this aspect becomes the core concept of the above formalism. Naturally, this results in a clear focus on test *validity*, which is defined purely on semantic grounds. In contrast to this, classical validity theory usually draws on statistical (usually correlational) analyses, which are blind for semantic relations.

The approach outlined in this paper sees itself not as opposed to psychometric theory but tries to complement it with an alternative way to conceptualize psychological tests. This can be especially frugal, if the theoretical implications of latent variable theory do not apply to test, or have shown to be incorrect. In some diagnostic settings the theoretical background of a test may even be irrelevant to the interpretation of test scores.

The main difference between classical psychometrics and the approach at hand, however, consists in the interpretation of construct scores obtained by a test. Whereas psychometrics is deeply connected with the theory of latent variables, the semantic approach makes no such assumptions whatsoever. Hence, tests are not interpreted as measurement devices, but as standardized assessment tools, which give empirical evidence on the correct use of an abstract term to describe a person. This is not only consistent with the widely accepted interpretation of psychological constructs as abstract concepts, but also avoids misleading analogies to physical measurements, which provoke common misinterpretations of psychological test results as 'objective measures'. Instead, the semantic approach to psychological tests presented in this paper explicitly models the definitional character of psychological constructs, resulting in a formal framework which is both theoretically sound and easy to apply in practice.

References

Alsina, C. (1985). On a family of connectives for fuzzy sets. *Fuzzy Sets and Systems, 16*(3), 231-235.

Belohlavek, R., & Klir, G. J. (2011). *Concepts and Fuzzy Logic.* Cambridge, London: MIT Press.

Blumer, H. (1940). The problem of the concept in social psychology. *American Journal of Sociology, 45*, S. 707-719.

Borkenau, P. (1988). The multiple classification of acts and the Big Five factors of personality. *Journal of Research in Personality, 22*, pp. 337-352.

Borsboom, D. (2005). *Measuring the mind. Conceptual issues in contemporary psychometrics.* Cambridge: Cambridge University Press.

Borsboom, D. (2008). Latent variable theory. *Measurement, 6*, 25-53.

Borsboom, D., & Cramer, A. O. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology, 9*, S. 91-121.

Borsboom, D., Mellenbergh, G., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203.

Buntins. (2014). *Psychologische Tests und mehrwertige Logik - Ein alternativer Ansatz zur Quantifizierung psychologischer Konstrukte.* Wiesbaden: Springer VS.

Buss, D. M. (1983). Act Prediction and the Conceptual Analysis of Personality Scales: Indices of act Density, Bipolarity, and Extensity. *Journal of Personality and Social Psychology, 45*, 1081-1095.

Buss, D. M., & Craik, K. H. (1980). The frequency concept of disposition: Dominance and prototypically dominant acts. *Journal of Personality, 48*, 379-392.

Buss, D. M., & Craik, K. H. (1981). The act frequency analysis of interpersonal dispositions: Aloofness, gregariousness, dominance and submissiveness. *Journal of Personality, 49*, 175-192.

Cramer, A. O., Waldorp, L. J., van der Maas, H. L., & Borsboom, D. (2010). Comorbidity: a network perspective. *Behavioral and Brain Sciences, 33*(2-3), S. 137-150.

Edwards, J. R., & Bagozzi, R. P. (2000). On the Nature and Direction of Relationships Between Constructs and Measures. *Psychological Methods, 5*(2), S. 155-174.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists.* New York: Erlbaum.

Ginsberg, A. (1954). Hypothetical constructs and intervening variables. *Psychological review, 61*(2), 119.

Gottwald, S. (2010). Many-Valued Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2010 ed.).

Gottwald, S., & Hájek, P. (2005). Triangular norm-based mathematical fuzzy logics. In E. Klement, & R. Mesiar (Eds.), *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms* (pp. 257-299). Amsterdam: Elsevier.

Gottwald, S., & Novák, V. (2000). Approach towards consistency of fuzzy theories. *International Journal of General Systems, 29*(4), 499-510.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Hájek, P. (1998a). Basic fuzzy logic and BL-algebras. *Soft Computing, 2*, 124-128.

Hájek, P. (1998b). *Metamathematics of Fuzzy Logic.* Dordrecht, Boston, London: Kluwer Academic Publishers.

Hájek, P. (2006). What is mathematical fuzzy logic. *Fuzzy Sets and Systems, 157*(5), 597-603.

Kline, P. (1998). *The new psychometrics. Science, psychology and measurement.* London: Routledge.

Kline, P. (2000). *The handbook of psychological testing, second edition.* London, New York: Routledge.

Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic.* New Jersey: Prentice Hall.

Kyngdon, A. (2008). Conjoint measurement, error and the Rasch model. *Theory and Psychology, 18*(1), 125-131.

Kyngdon, A. (2008). The Rasch model from the perspective of the representational theory of measurement. *Theory and Psychology, 18*(1), 89-109.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental scores.* Reading, MA: Addison-Wesley.

Łukasiewicz, J. (1920). O logice trojwartosciowej. *Ruch Filozoficny, 5*, 170-171.

Łukasiewicz, J. (1970). On three-valued logic. In L. Borkowski (Ed.), *Jan Łukasiewicz. Selected Works.* Amsterdam: North-Holland.

MacCorquodale, K., & Meehl, P. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*(2), 95.

Maraun, M. D., & Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas in Psychology, 31*(1), S. 32-42.

Markus, K. A., & Borsboom, D. (2013). Reflective measurement models, behavior domains, and common causes. *New Ideas in Psychology, 31*(1), S. 54-64.

Micallef, L., & Rodgers, P. (2014). eulerAPE: Drawing Area-proportional 3-Venn Diagrams Using Ellipses. *PLoS ONE, 9*(7).

Michell, J. (1990). *An introduction to the logic of psychological measurement.* Hillsdale: Lawrence Erlbaum Associates.

Michell, J. (1999). *Measurement in psychology: Critical history of a methodological concept.* Cambridge: Cambridge University Press.

Michell, J. (2013). Constructs, inferences, and mental measurement. *New Ideas in Psychology, 31*(1), 13-21.

Mirels, H. (1982). The illusory nature of implicit personality theory. *Journal of Personality, 50*, pp. 203-222.

Novák, V. (2012). Reasoning about mathematical fuzzy logic and its future. *Fuzzy Sets and Systems, 192*(1), 25-44.

Post, E. L. (1921). Introduction to a general theory of elementary propositions. *American Journal of Mathematics, 43*, 163-185.

Priest, G. (2008). *An Introduction to Non-Classical Logic. From if to is. 2nd edition.* Cambridge: Cambridge University Press.

Russell, J. A., & Fehr, B. (1994). Fuzzy Concepts in a Fuzzy Hierarchy: Varieties of Anger. *Journal of Personality and Social Psychology, 67*(2), 186-205.

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology, 31*(1), S. 43-53.

Slaney, K., & Racine, T. (2013). What's in a name? Psychology's ever evasive construct. *New Ideas in Psychology, 31*(1), 4-12.

Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining.* Boston: Pearson Addison Wesley.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353.

Zadeh, L. A. (1975a). The concept of a linguistic variable and its application to approximate reasoning i. *Information Sciences, 8*(3), S. 199-251.

Zadeh, L. A. (1975b). The concept of a linguistic variable and its application to approximate reasoning ii. *Information Science, 8*(4), S. 301-357.

Zadeh, L. A. (1975c). The concept of a linguistic variable and its application to approximate reasoning iii. *Information Science, 9*(1), S. 43-80.