Comments on "Towards a Framework for Software Measurement Validation"

Sandro Morasca, *Member*, *IEEE Computer Society*,
Lionel C. Briand, Victor R. Basili, *Fellow*, *IEEE*,
Elaine J. Weyuker, Member, *IEEE Computer Society*,
and Marvin V. Zelkowitz, *Senior Member*, *IEEE*

Abstract—A view of software measurement that disagrees with the model presented in a recent paper by Kitchenham, Pfleeger, and Fenton, is given. Whereas Kitchenham, Pfleeger, and Fenton argue that properties used to define measures should not constrain the scale type of measures, we contend that that is an inappropriate restriction. In addition, a misinterpretation of Weyuker's properties is noted.

Index Terms—Software measurement, measurement theory, measurement scales, axiomatic approaches, software complexity properties.

1 Introduction

Kitchenham, Pfleeger, and Fenton [3] questioned the way properties have been used in the literature to assess software measures. We have two main comments on their criticisms.

2 Attribute Properties and Measurement Scales

First, the authors propose that properties that imply or exclude any particular measurement scale in the definition of a measure cannot be used. This is stated clearly in the following paragraph ([3] p. 932, last paragraph):

> "2) Since an attribute may be measured in many different ways, attributes are independent of the unit used to measure them. Thus, any definition of an attribute that implies a particular measurement scale is invalid. Furthermore, any property of an attribute that is asserted to be a *general* property but implies a specific measurement scale must also be invalid."

Adherence to this model will seriously impede the appropriate definition of attributes, particularly when there is a well understood intuition. There is no problem defining properties that at least permit ordering over the set of entities. Without such properties, we end up abstracting away all relevant structure from our model, limiting our ability to say anything of interest. It is true that such properties would only be relevant for measures of the attribute that are defined on an ordinal scale or higher. Nevertheless, this does not make such properties invalid, as stated in [3]. Experimental physics has successfully relied on attributes such as temperature that imply measurement scales in the definition of their properties. Other examples will help clarify our point.

- *S. Morasca is with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy. E-mail: morasca@elet.polimi.it.*
- *L.C. Briand is with the Fraunhofer Institute for Experimental Software Engineering (FhG IESE), Sauerwiesen 6, D-676761, Kaiserslautern, Germany. E-mail: briand@iese.fhg.de.*
- *V.R. Basili and M.V. Zelkowitz are with the Computer Science Department, University of Maryland, College Park, MD 20742. E-mail: {basili, mvz}@cs.umd.edu.*
- *E J. Weyuker is with AT&T Research–Labs, Room 2A-429, 600-700 Mountain Ave., P.O. Box 636, Murray Hill, NJ 07974. E-mail: weyuker@research.att.com.*

Manuscript received Aug. 12, 1996.
Recommended for acceptance by D.R. Jeffery.
For information on obtaining reprints of this article, please send e-mail to: transse@computer.org, and reference IEEECS Log Number S96132.

Consider the notion of the size of an object. Our intuitive understanding about the concept of size is that when one "puts together" two objects O_1 and O_2 to obtain a third object O_3, then the size of O_3 is not smaller than the size of O_1 or O_2. (Although the operation of "putting together" may be formally defined, for brevity's sake, we do not provide such a definition here.) Since this simple property is not appropriate for nominal scale size measures, it would be considered invalid for any size attribute, according to [3]. However, the usual understanding of the attribute size can be formalized through properties of size measures requiring at least the possibility of comparing objects' sizes. Our understanding of the attribute size may go even further. In fact, in [2], we propose simple and widely-acceptable properties that imply the ratio scale.

As a second example, consider the notion of distance between two elements of a set S. According to the standard definition, distance is defined as a function:

$$d: \quad S \times S \rightarrow R_+$$

(R_+ is the set of nonnegative real numbers) that satisfies the following three axioms:

Axiom 1.	$\forall x, y \in S$	$(d(x, y) \geq 0$ **and** $(d(x, y) = 0 \Leftrightarrow x = y))$
Axiom 2.	$\forall x, y \in S$	$(d(x, y) = d(y, x))$
Axiom 3.	$\forall x, y, z \in S$	$(d(x, y) \leq d(x, z) + d(z, y))$

These axioms exclude

- nominal scales, since they contain the "≤" operator;
- ordinal scales, since they contain the "+" operator;
- interval scales, since Axiom 3's truth value is not invariant under the admissible transformation for interval scales, i.e., Axiom 3 does not imply the following formula (R is the set of real numbers):

$$\forall x, y, z \in S, \forall a > 0, \forall b \in R\ (a \cdot d(x, y) + b \leq a \cdot d(x, z) + b + a \cdot d(z, y) + b)$$

Axiom 3's truth value is invariant under the admissible transformation for the ratio scale, i.e., Axiom 3 implies the following formula

$$\forall x, y, z \in S, \forall a > 0\ (a \cdot d(x, y) \leq a \cdot d(x, z) + a \cdot d(z, y)).$$

Therefore, Axiom 3 implies the ratio scale, and hence, according to [3] the three axioms usually provided for distance are invalid. This view of meaurement is so narrow and restrictive that it limits our ability to define properties that adequately characterize attributes, even for very well-understood attributes.

We therefore conclude that acceptance of the perspective proposed in [3] has important consequences, including:

1) If we discard some properties, we may be discarding a good deal of relevant information about the attribute. Therefore, our modeling of the attribute will not be as accurate as it could be.
2) If we discard some properties, we will have a less powerful mechanism for checking whether a function that is proposed as a measure for an attribute actually is a measure for that attribute.

It is certainly not true that all attributes can be appropriately defined by properties that imply the ratio, interval, or even ordinal scale (e.g., the color of a physical object). However, as argued above, this does not imply that we should forbid any attributes from being defined by properties that do imply a particular measurement scale or prevent some measurement scales.

Although some attributes used in software engineering (e.g., complexity, cohesion, coupling) are not as well-understood as distance or size, it does not follow that we should prohibit the use of properties that constrain the scale type of a measure. Indeed, an important purpose of using properties as a means of defining measures is to help codify intuition and make underlying assumptions explicit. In fact that is exactly why Euclid introduced the axiomatic method for geometry more than 2,000 years ago.

3 Weyuker's Properties

Another point of this paper involves criticisms of the properties Weyuker proposed in [4]. First, the authors repeat Zuse's statement [5] that Weyuker's axioms are inconsistent from a Measurement Theory point of view ([3] p. 932, last paragraph):

> "Thus, while Zuse criticises Weyuker's complexity measure properties as contradictory because one (property 5) implies a ratio scale and another (property 6) explicitly excludes a ratio scale …"

They describe Weyuker's properties as "disputed" and "caution researchers to avoid justifying measures on the basis of either disputed properties or …" As argued in [1], a careful reading of Zuse's book demonstrates that Zuse's criticisms are unfounded. Concisely, in [1] we show that Zuse's criticisms only prove that Weyuker's properties are not compatible with the fact that the underlying empirical system of a measure assumes an extensive structure. However, the fact that the underlying empirical system of a measure assumes an extensive structure is a sufficient condition to obtain a ratio scale measure, but is by no means a necessary one. Although Zuse refers to Weyuker's properties as contradictory, they are not contradictory in the usual mathematical sense of being incapable of being satisfied at the same time. Some of the properties do require the ratio scale, but there is nothing inappropriate about this.

In addition to Zuse's criticism, another erroneous criticism is introduced in [3], p. 939.

> "3) *Each unit of an attribute contributing to a valid measure is equivalent*. This seems to be standard measurement practice. Weyuker's property 7 relates to this issue. She, in fact, asserts the *converse* of this assumption by claiming that program complexity should be responsive to the order of statements in a program. It seems here that Weyuker is confusing the attributes program correctness and/or psychological complexity with structural complexity. It is unlikely that a random re-ordering of program will be correct or understandable, but a re-ordering would not necessarily be more structurally complex."

Weyuker's property 7 asserts that there exist two programs P and Q, where Q is a re-ordering of P, such that the complexity of P is *different* from the complexity of Q [4]. This property does not assert that, by re-ordering a program, one obtains a new program which would necessarily be *more* or *less* structurally complex than the original one. Weyuker's property 7 states that program complexity may be responsive to the order of statements. It does not contradict the statement made by Kitchenham, Pfleeger, and Fenton:

> "a re-ordering would not necessarily be more structurally complex."

Just as Zuse's criticism in [5] with respect to Axioms 6, 7, and 9 was caused by a misinterpretation or misrepresentation, Kitchenham, Pfleeger, and Fenton have misinterpreted Weyuker's axiom 7.

References

[1] L.C. Briand, K. El Emam, and S. Morasca, "On the Application of Measurement Theory in Software Engineering," *Empirical Software Eng.: An Int'l J.*, vol. 1, no. 1, pp. 61-88, 1996.
[2] L.C. Briand, S. Morasca, and V.R. Basili, "Property Based Software Engineering Measurement," *IEEE Trans. Software Eng.*, vol. 22, no. 1, pp. 68–86, Jan. 1996.
[3] B. Kitchenham, S.L. Pfleeger, and N. Fenton, "Towards a Framework for Software Validation Measures," *IEEE Trans. Software Eng.*, vol. 21, no. 12, pp. 929–944, Dec. 1995.
[4] E.J. Weyuker: "Evaluating Software Complexity Measures," *IEEE Trans. on Software Eng.*, vol. 14, no. 9, pp. 1,357–1,365, Sept. 1988.
[5] H. Zuse, *Software Complexity: Measures and Methods*. De Gruyter, 1991.