On Conceptual Struggles over "Testing"

Krzysztof M. Brzeziński
Institute of Telecommunications
Warsaw University of Technology
Email: kb@tele.pw.edu.pl

***Abstract*—Testing, one of the key activities in engineering, is full of preconceptions. The general concept of testing is not clear, both for professionals and for educators, as "testology" has not established itself as a science yet. It is thus possible that, at the university level, the wrong concept of testing is being taught. We identify the main preconceptions (including the one that conflates testing with active testing), and propose ways to avoid implanting such preconceptions in curricula.**

I. Introduction

Testing is one of the key activities in engineering. The direct context of this paper is behavioural testing of ICT (Information and Communications Technology) systems characteristic of telecommunications, industrial control, and computing in general. Clearly, testing in this particular context should be recognizable as a specialized embodiment of the "general concept(s) of testing". Surprisingly, this does not seem to be the case – there is currently no recognized scientific discipline that would be held responsible for developing and fostering such concept(s). We claim, however, that such science-in-the-making is currently emerging. As it seems to follow the past development path of *metrology* [1], we will refer to this still unnamed future science[1] as *testology*.

According to its role in engineering, testing should be practised professionally, which also requires that *teaching* be done at the appropriate level of sophistication and with proper goals [4]. Disfunctional teaching (e.g., putting emphasis on an irrelevant or misleading facet "because the theory is so well developed" [4], and thus "easier" on the educators) may implant in future professionals the undesirable thinking patterns that may be hard to uproot. Preconceptions seeded in academia tend to strongly petrify the discipline, as the language and the whole conceptual setting is then "officially" defined, and it is not uncommon for professionals to keep and refer to textbooks and notes from their academic studies.

This vicious circle has been noted and addressed in many publications, although mainly in relation to coarser areas, such as electrical engineering and computer science. The inadequacies identified so far pertain, e.g., to "fundamental skills and knowledge" addressed by teaching [4], and to the use of formal (i.e., based on syntactic structures) modelling as a criterion for the university level [3]. These reservations are also directly applicable to testing. For example, "*it is natural that formal methods should underlie principled testing methods*" [5], and formal model-based testing [6] is indeed gaining recognition, but these formal aspects are still not given due consideration and coverage in university courses on testing. We, however, go one step deeper, and address mainly the *conceptual framework* of testing, which is mostly *pre-formal*: formalization may help with relating concepts and gaining better understanding of the consequences of particular conceptual choices, but it is important to know *what* is being formalized. We thus ask, and partly answer, two quite fundamental questions: (a) whether we are, after all, using and teaching the right *concept(s)* of testing, and (b) how the current status of testology bears on the possible answers to (a).

Assuming that an important part of an academic curriculum should be focused on a set of *core concepts*, or "first principles" [2] of testing, we show that what is currently believed to be, and taught as, this *core* is fundamentally flawed by preconceptions. A part of these preconceptions, concerning the *active* character of testing, is addressed in more detail.

II. Testology

There is an identifiable research activity that deals specifically with testing, herein referred to as "testology". It has its own research subject, a body of researchers, accepted theoretical results with repeatable practical applications, and publications with "testing" as the principal keyword. However, it has also peculiar features that make it rather "special".

A. The Disciplinary Identity of Testology

Testing (as testing *of something*) is interdisciplinary: it accompanies many fields of human activity. It is thus tempting to consider testing as a sub-problem, whose methodology belongs with a given field, forms a subset of the innate methodology of that field, and also inherits its particular language, as in Wittgenstein's *language games* (patterns of understanding as language-related *habits*). This approach has resulted in a number of different schools of thought on testing, with their terminological, and also quite fundamental *conceptual* flavours, and with very little tendency to actively seek interactions with peer communities [7]. In this aspect, clear analogies can be drawn to metrology. Its individual incarnations (mechanical; electrical, chemical metrology, etc.) were being developed quite independently. Currently, however, its conceptual and terminological framework is relatively stable and widely accepted *across* its application domains [1].

This research was conducted in the framework of a research project contracted by the Polish Ministry for Science and Higher Education, and was financed with the 2007-2010 research funds.

[1]The traditional distinction between *sciences* and *engineering* disciplines has been found not epistemologically useful at the university level [2], [3].

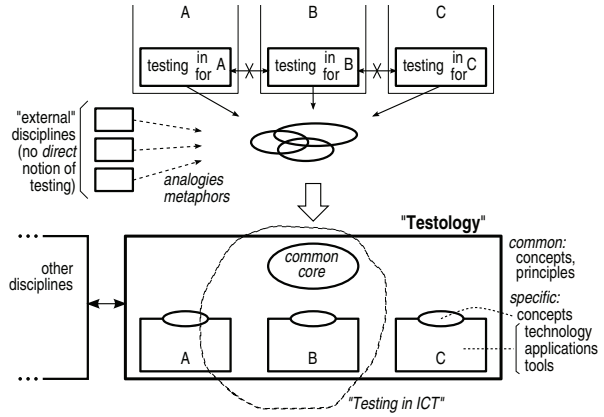

Figure 1. "Testology" – current state and postulated development

On the other hand, where an attempt to capture the overall framework of testing is undertaken, it is usually approached by enumerating its well understood flavours, or by presenting one of many frameworks *as if* it was *the* most general framework.

It is a popular idea that any science is built upon *definite basic concepts*. According to the alternative view, "*the definition and systematization of the fundamental notions of a science are a* late *stage in its development*" [8]. Another, orthogonal dimension of the ongoing debate on disciplinary identity and academic legitimacy [9] identifies two positions: according to one, a well-defined theoretical core is a "*necessary condition for having a clear disciplinary identity*", while the other points to difficulties in defining what a "theoretical core" is, and advocates more diversity. Whatever the case may be, it is clear that the set of basic concepts of testology is not complete, and yet presenting and discussing this set is an important, indispensable "*scientific basis proper for the university level*" of education [3]. This apparent contradiction seems to be *the* main methodological concern. Even in more mature fields, which enjoy a seemingly complete, but often *impractically large* conceptual framework, there is a tendency to use a certain dangerous educational "trick", described in [4]: "*Traditional engineering courses often try to circumscribe a body of material, creating a complete and coherent subject with consistent notation, and within which every problem has a right solution. . . . when related to the real world this approach is seriously misleading*". In case of a science *in statu nascendi*, like testology, this "trick" is particularly counterproductive. Here, the conceptual framework must be taught while it is being developed, without cutting off any possible future paths of its development. These concerns are illustrated in fig.1.

B. Patterns and Metaphors

The ability to identify problems as instances of already established *patterns* is a part of scientific competence. In education, both the results of such pattern-matching, and the skills for so doing are important and should be taught. When no suitable pattern can be recognized (not only for individual research problems, but also for the elements of a conceptual framework itself), an in-house solution must be developed, or an *external* pattern should be found and adopted. The popular "donors" are ontology, epistemology, and semiotics [9], all situated within the "umbrella science" of philosophy. The educational idea postulated herein is not to hide, but to *expose* the already identified and possible future conceptual links of testology with these and other disciplines. Borrowing concepts is bidirectional, and *both* directions are worth making explicit. Testology, not enjoying the status of a university discipline, will probably *not* be considered as a viable donor of concepts (archetypes, design paradigms, etc.) for other, "external" disciplines that may be faced with problems that *could* be mapped onto the already solved, or at least formulated, testing problems (as in [10], where applicability of testing concepts to *intrusion detection* was established). Showing how testology could be *useful* as a donor is one of the proposed priorities.

The "borrowing" is likely to take on the form of a *conceptual metaphor*, for "*conceiving of one thing* [a target] *in terms of another* [a source]" [11]. Apart from general human experience (as in: "*Life* [a target] *is a journey* [a source]"), conceptual metaphors also function in *teaching*, to illustrate and induce the understanding of concepts already understood *by others*. In case of testology the application of this kind of metaphors is limited, as the pre-existing understanding is also limited, and possibly misleading. The third broad application of metaphors, of direct relevance to our concerns, is in scientific research. An example (apparently not explored before) of such metaphor might be "*Testing* [a target] *is a measurement* [a source]". Testing may then be conceived of as a particular kind of measurement – this metaphorical iconic relation (partial structural resemblance) to measurement may serve as a strong evocative (insight-inducing) tool [9].

C. A Meta-Linguistic Perspective

A plausible reason for the failure to recognize patterns, even if they *are* actively sought, is the inability or reluctance to transcend a community-specific *language-game*. According to the philosophical and linguistic idea captured by the Sapir-Whorf hypothesis [12], language puts limits to what a community *can* say, and what it finds *worth* discussing, because it influences (or, in the most orthodox version, predetermines) how "the world" is perceived. Also in technical contexts there is no reason for dismissing such concerns. When a pattern is formulated in a "foreign language", it may not be understood and appreciated. This phenomenon was identified in Computer Science by Lamport [13], who even talks of the "Whorfian syndrome" [2010, private communication]. The same can be stated of testology – its individual branches have different language-induced "world-views". It is not easy, however, to change the language of a scientific community, and thus – its perspective on the field of research. Doing this *ex caethedra* does not seem *effective*. To counter this obstacle, a meta-linguistic approach was proposed in [14]. In this approach, various community-specific ways of talking about "testing-like" concepts and practices are investigated, a harmonized concept of testing is distilled from what these communit-

ies talk *about*, and spurious terminological differences and incompatibilities are identified and dismissed. To illustrate the problem, consider various euphemisms that are used by different communities to avoid the "politically incorrect" term "passive tester": *observer*, *trace checker*, *the oracle*, *passive monitor* (sic!), *arbiter*, *supervisor*.

III. Challenging the Scope of "Testing"

To illustrate the troubles with the concept of testing, we now concentrate on establishing whether "passive testing is testing". In telecommunications, computer science, and many other disciplines, testing is tacitly understood, explicitly defined, and also *taught* as essentially *active*: a tester is said to stimulate, control, and enforce. Apparently, if testology were Kuhn's *normal science*, "being active" would be a crucial element of its paradigm. On the other hand, for almost 30 years there has been ongoing work on the non-paradigmatic, and thus *niche* research subject of *passive testing*, but it is still considered by many authors, including Tretmans [15], to be a mere *façon de parler*, the case of confusion of tongues. The non-mainstream position of such reseach makes it unlikely for other disciplines to look into it for useful testing metaphors. Consequently, "passive testing" is almost banned in teaching, for the sake of the "purity" of the conveyed concept, and any possible analogies and metaphors are not brought into the attention of students.

Testing is usually understood as an activity, in which a tester (*a*) *generates* (applies) stimuli that provoke or stimulate phenomena to be investigated; (*b*) *observes* phenomena as they appear under the influence of applied stimuli; (*c*) *analyses* the relation between the applied stimuli and observed phenomena; and (*d*) basing on these checks, *decides* on a suitable verdict. This understanding, which clearly stipulates the *active* technique, is often taken as a definition of testing (which also filters out the views that are *not* to be taught in testing classes). The starting point, however, can be chosen differently. In the most general sense, testing may be understood as [14]: (*a*) an activity with at least some empirical, *experimental* elements, which results can only be established *a posteriori*; (*b*) where experiments are conducted on a particular object – *Thing under test* (**Tut**); (*c*) leading to the evaluation of a certain entity that partakes in testing – the *Object of assessment* (**Ooa**), which may, or may not be a **Tut**; and (*d*) aimed at a certain *goal*. This goal may be expressed using different languages. In the language of the formal testing community [16], the goal is to establish whether a given *relation* (equivalence or preorder) holds between a **Tut** and a given *reference* **Ref**. In the language of the Scientific Method, also adopted by the "logical" school of testing and implicitly used in the industry-oriented testing framework described in [17], the goal is to establish whether a given *hypothesis* (which also means – all its necessary consequences, or "requirements") concerning a **Tut** can be regarded as true. In the language of epistemology, the goal is to obtain *knowledge* (which can be taken to mean: *justified true belief*) whether **Tut** is correct. Note that each of these linguistic settings brings its own insights into the notion of testing.

A *matrix of choices* concerning the roles assigned by different schools to different elements of the conceptual space of testing is presented in [10]. Observations contained there show how *arbitrary* the ICT community's choices and commitments w.r.t. the concept of testing are, and how much flexibility is to be gained by surrounding the petrified choices ("active" being but one of them).

Nowhere in the exposition presented above "being active" appears as a necessary property of testing. In Aristotelian theory of predication [18], such property might be *essential* (explicitly included in the definition of testing), or might be a *proprium* – still necessary, and logically deriveable from a definition, but not explicitly present in this definition. As it is, "being active" appears to be rather an *accident* of testing.

The important pre-technical, philosophical (epistemological) aspects of the concept of testing are present in the *Scientific Method* (SM) – a paradigm of sound scientific enquiry, with its phase of "testing a hypothesis" – performing *experiments* aimed specifically at confirming or denying the existence of predicted phenomena. The "unusual" context of the Scientific Method can be brought closer to what is customary in ICT-related testing, by stating "**Tut** is correct" as a hypothesis, and accepting, as the logical consequences of this hypothesis, the *requirements* or elements of a *model*.

Within the Scientific Method, "experiment" and "test" have, for all practical purposes, the same sense [19], and an experiment is not required to be necessarily *active*, i.e., that in which influence is purposefully exerted upon investigated phenomena. John Stuart Mill distinguishes two kinds of experiments: *pure observation* and *artificial experiments* [20], and finds place for *both* in organized scientific enquiry. There is thus nothing unusual or "politically incorrect" in considering *passive tests*, based on quasi-experiments, i.e., natural, non-controlled experiments. They involve the observation and assessment of phenomena that are not invoked (provoked, stimulated, influenced) by a tester. This lack of influence may be intended or required due to: (*a*) the nature of a phenomenon, which does *not allow* for such influence; (*b*) a phenomenon being "intensive enough" so that any stimulus is actually *not needed*; (*c*) the properties of a phenomenon itself – there may be reasons for not allowing *any* stimulation whatsoever, in order not to distort it; (*d*) the properties of a system in which investigated phenomena occur, whose *integrity, safety, and performance* may critically depend on non-interference with its internal parts and processes; and (*e*) technical issues – it may be too inconvenient or costly to equip a tester with a "sender". Interesting thoughts on "testing", which reveal both its active and passive connotations, may also be obtained by consulting the regular, dictionary entries for this term.

One of the ways to converge on the concept of testing is to consider *monitoring*, and in particular – to realize *how* (in which capacity) monitoring is different from testing (*any* kind of testing), as the common tendency so far has been to conflate monitoring with passive testing. The initial approach is based

on differentiating the *level* of interpretation. *Monitoring* is a technical counterpart of *watching* a particular aspect of a system, and provides a stream of (syntactically processed, e.g., filtered, and suitably presented) data on the behaviour of an investigated system, intended to be interpreted (*seen*) by an external process, which may also be a human observer. Unlike monitoring, *passive testing* involves both, the (syntactic) process of *watching*, and the (semantic) interpretative process of *seeing*, and provides a stream of *interpretations*, or *valuations*, of monitored data.

It is, however, not uncommon to see some semantic interpretations (e.g., raising an alarm if a threshold is exceeded) shifted from an external interpretive process to the monitoring process itself, which yields *extended monitoring*. Even if several layers of such interpretations are present, there is no clear point at which (extended) monitoring would necessarily, and "magically", change into passive testing. Here the initial approach fails – the output of *both* processes is a stream of interpretations. The second, complementary approach is thus based on the aim, or "direction", of the interpretation [21]. Let a pair: $< B, C >$ consist of a particular *behaviour*, and circumstances (*conditions*) in which it is exhibited. Monitoring and (passive) testing differ in the logical ordering of the elements of this pair. In (extended) monitoring, interpretation serves to infer, from the observed behaviour, its circumstances, or the general *mode of operation* of a system (such as "being overloaded" or "being under attack"). In testing, interpretation is related to pre-defined conditions, called in this context *test purposes*. In active testing, a test system, steered by a test program, *establishes* (forces) these conditions, while in passive testing a test system *recognizes* them. We may conclude that, by jointly using both approaches together, passive testing *can* be distinguished from monitoring.

IV. Concluding Remarks

Admittedly, there are sources of "standardized knowledge and skills" concerned with testing in general, and testing in software-dominated systems in particular [22], [17], [23]. They form the core of curricula for vocational study that leads to obtaining the diploma of a "certified tester" (such as those granted by iSTQB – International Software Testing Qualifications Board [24]). There is noticeable pressure to focus higher education on vocational skills, but "no set of vocational skills has any significant longevity" [4], so such sources of "canned knowledge" are ill suited for university-level courses.

Our suggested approach to teaching "testing" at the university level is consistent with the theses and guidelines presented in [4], where it is strongly suggested to "*focus on* fundamentals *that will serve the students well for an entire career*", to "*look for* fundamental *skills and knowledge*", to "*prepare students to select abstractions, not just technologies*", and to allow for "*more room ... for nontechnical subjects such as economics and humanities*". Similarly, we corroborate the observations made in [2] that it is possible to "*learn a lot from experience, history, and* philosophy *of other disciplines and associated images of science*".

Concerning the current set of preconceptions about testing (including the bias towards "active testing"), an anecdotal solution may be proposed, due to Max Planck (cited in [3]): "*not to try to convince the opponents, but to wait until they gradually die out*". While executing this advice, it is expedient to cater for the *right* education of the next generation.

References

[1] J. Jaworski, R. Morawski, and J. Olędzki, *Wstęp do metrologii i techniki eksperymentu*. WNT, 1992.

[2] H. Akkermans and J. Gordijn, "What is this Science called Requirements Engineering?" in *Proc. 14th IEEE Int. Conf. Requirements Engineering*, 2006, pp. 273–278.

[3] R. Boute, "Teaching and Practicing Computer Science at the University Level," *SIGCSE Bulletin of Symbolic Logic*, vol. 41, no. 2, pp. 24–30, 2009.

[4] E. A. Lee and D. G. Messerschmitt, "Engineering an Education for the Future," *Computer*, pp. 77–85, Jan. 1998.

[5] J. Woodcock, P. G. Larsen, J. Bicarregui, and J. Fitzgerald, "Formal Methods: Practice and Experience," *ACM Computing Surveys*, vol. 41, pp. 1–36, 2009.

[6] M. Broy, B. Jonsson, J.-P. Katoen, M. Leucker, and A. Pretschner, Eds., *Model-Based Testing of Reactive Systems*, ser. Lecture Notes in Computer Science, vol. 3472. Springer, 2005.

[7] R. M. Hierons, K. Bogdanov, J. P. Bowen, R. Cleaveland, J. Derrick, J. Dick, M. Gheorghe, M. Harman, K. Kapoor, P. Krause, G. Luettgen, A. J. H. Simons, S. Vilkomir, M. R. Woodward, and H. Zedan, "Using formal specifications to support testing," *ACM Computing Surveys*, 2008.

[8] C. J. Ducasse, *Philosophy as a Science, its Matter and its Method*. New York: O. Piest, 1941. [Online]. Available: http://www.ditext.com/ducasse/ducasse.html

[9] K. M. Brzeziński, "Standards are Signs," in *15th EURAS Annual Standardization Conference 'Service Standardization'*, J.-C. Graz and K. Jakobs, Eds. EURAS, 2010, pp. 43–60.

[10] ——, "On Common Meta-Linguistic Aspects of Intrusion Detection and Testing," *Int. J. Information Assurance and Security (JIAS)*, vol. 2, no. 3, pp. 167–178, 2007.

[11] G. Lakoff and M. Johnson, *Metaphors we live by*. The University of Chicago Press, 1980.

[12] R. N. St.Clair, "Cultural Wisdom, Communication Theory, and the Metaphor of Resonance," *Intercultural Communication Studies*, vol. VIII-1, pp. 79–103, 1998-9.

[13] L. Lamport, "Computation and State Machines," 2008. [Online]. Available: http://research.microsoft.com/en-us/um/people/lamport/pubs/state-machine.pdf

[14] K. M. Brzeziński, "Towards the Methodological Harmonization of Passive Testing Across ICT Communities," in *Engineering the Computer Science and IT*, S. Soomro, Ed. In-Tech, 2009, ch. 9, pp. 143–168.

[15] J. Tretmans, "Testing Concurrent Systems: A Formal Approach," in *CONCUR*, 1999, pp. 46–65.

[16] S. D. Bruda, "Preorder Relations," in *Model-Based Testing of Reactive Systems*, ser. LNCS, M. Broy, B. Jonsson, J.-P. Katoen, M. Leucker, and A. Pretschner, Eds., vol. 3472, 2004, pp. 117–149.

[17] ISO/IEC 9646, *Conformance Testing Methodology and Framework*. ISO/IEC, n.d., vol. 1–7.

[18] H. W. B. Joseph, *An Introduction to Logic*, 2nd ed. Oxford University Press, 1925.

[19] M. Heidelberger, "Experimentation and Instrumentation," in *Encyclopedia of Philosophy*, D. M. Borchert, Ed. Thomson Gale, 2006, vol. 10: Appendix, pp. 12–20.

[20] J. S. Mill, *Of Observation and Experiment*. Routledge and Kegan Paul, 1974, ch. VII.

[21] K. M. Brzeziński, "Towards Practical Passive Testing," in *Parallel and Distributed Computing and Networks (PDCN'05)*, 2005, pp. 177–183.

[22] IEEE Std 1044, *IEEE Standard Classification for Software Anomalies – Description*. IEEE, 1993.

[23] ISTQB, *Standard Glossary of Terms Used in Software Testing, version 2.0*. ISTQB (Glossary Working Party), 2007.

[24] ——, *Certified Tester. Advanced Level Syllabus*. ISTQB, 2007.