Preparing Meta-Analysis of Metamodel Understandability

Susanne Patig[1]

[1] University of Bern, IWI, Engehaldenstrasse 8, CH-3012 Bern, Switzerland
susanne.patig@iwi.unibe.ch

Abstract. Metamodels are designed to be used by machines and humans. For human users, the understandability of the metamodel is important. Experimental investigations of understandability in computer science have led to conflicting results. To resolve such conflicts and gain insights into the nature of some phenomenon beyond singular experiments, meta-analysis can be applied, i.e., the statistical analysis of results obtained by other (primary) empirical studies. This paper shows the current obstacles for a meta-analysis of metamodel understandability: They consist in the heterogeneity of the individual experiments and deficient reporting. The paper provides a framework to increase the comparability of experiments on understandability. Such comparability enables future meta-analysis.

Keywords: Understandability, Metamodels, Experimental Research

1 Motivation

Designing and modifying metamodels are major topics of model-driven development. Metamodels must be understandable for both machine and human users. Following a definition of language understandability in cognitive psychology [1], the *understandability* of a metamodel means the effort required to read and correctly interpret its constructs and their connections. Understandability is a prerequisite both for reading artifacts (like documents or source code) that have been created by a applying a metamodel (*comprehension*) and for creating such artifacts (*specification*).

The ‘understandability’ of a metamodel for a machine shows up in error-free compilation. For human users, metamodel understandability must be empirically investigated, usually by controlled experiments. The results of such experiments are conflicting (see Section 3).

Conflicting empirical results can be statistically evaluated by meta-analysis (see Section 2). Meta-analysis could increase our knowledge about the nature of understandability – also to facilitate future metamodel design or modification. But, Section 3 shows that meta-analysis on metamodel understandability is currently hindered by (1) the heterogeneity of the conducted experiments and (2) insufficient reporting of the experimental results. This paper provides a framework to achieve comparability of experiments on metamodel understandability (see Section 4), which is a prerequisite for meta-analysis. Appropriate reporting guidelines exist (e.g., [16], [19]).

2 Meta-Analysis

Meta-Analysis is the statistical evaluation of numerical results that have been obtained by other (primary) studies [25], [13]. Hence, it is a kind of secondary research that aims at (1) finding evidence for some investigated phenomenon beyond individual studies (by calculating general descriptive statistics), (2) explaining conflicting results (by discovering new influencing variables), and (3) removing the bias potentially contained in 'normal' literature reviews of empirical studies [27].

Literature reviews often concentrate only on significant results that support the reviewer's theoretical position. But, statistical significance can be misleading, because it is affected by sample size [9]: If the same experiment is conducted independently, a larger sample may yield a statistically significant result, while a smaller one does not. The 'empirical truth' can be revealed by *effect size*. *Effect size* expresses the magnitude of a result, independently of sample size [9]. Table 1 summarizes main effect size measures and defines what constitutes a small, medium or large effect.

Table 1. Measuring Effect Size

Effect size measure	**Statistical test procedure**	**Refe-rence**	**Effect**		
			Small	**Medium**	**Large**
$d = \frac{\mu_1 - \mu_2}{\sigma}$ or $r_{es,t} = \sqrt{\frac{t^2}{t^2 + df}}$	t-test*	[10]: d	0.2	0.5	0.8
		[10]: r	0.1	0.3	0.5
$\omega = \sqrt{\frac{\chi^2}{N}}$	χ^2-test	[10]	0.1	0.3	0.5
$\eta^2 = \frac{\sigma_\mu^2}{\sigma^2}$	F-test (ANOVA)	[10], [9]	0.01	0.06	0.14
$r_{es,z} = \frac{\lvert z \rvert}{\sqrt{N}}$	U-test or any other that yields a z-score♦	[28]	0.1	0.3	0.5

μ_1, μ_2: Group means, σ: Standard deviation. σ^2 (σ_μ^2): Total (Between group means) variance, t, χ^2 (z): Test statistics (z: normal distribution), N: Total number of participants ($N = \Sigma n_i$), df: Degrees of freedom (df = n-2; n: [constant] number of participants of each group)

* Between-subjects design: σ of either group, within-subjects design: adjusted σ [9].
♦Requires $N \geq 25$ to calculate the z-scores by assuming normal distribution [10].

Meta-analysis typically integrates the effect sizes of singular studies. The *basic steps* are as follows [27], [25]:

1. Define the independent and the dependent variables of interest.
2. Systematically collect the studies to be included in the meta-analysis.
3. Estimate effect sizes for each study.
4. Combine the individual effect sizes to calculate and test the central tendency (e.g., the mean or median) and dispersion (e.g., variance) of the overall effect.

Various ways of combining effect sizes exist (see, e.g. [27]). The combined effect size quantifies the overall magnitude of some observed result, at least in the population of the included studies. To yield useful results from meta-analysis, the included studies must satisfy the following *requirements* [25]:

[RQ1] They must be of the same type (e.g., controlled experiments or case studies).

[RQ2] They must test the same hypothesis. Since a statistical hypothesis assumes that the independent variable(s) will *cause* the changes in the dependent variable(s) [28], these variables should be identical or comparable.

[RQ3] Often several measures for the same variable exist. Ideally, all included studies should use the same or comparable measures.

[RQ4] The studies should report effect sizes or provide at least statistics according to Table 1 or raw data to calculate the effect sizes.

The next section will show that current experiments on the understandability of metamodels do not satisfy these requirements.

3 Meta-Analysis of Research on Metamodel Understandability

This section sketches a failed attempt of meta-analysis – to prepare the ground for the framework in Section 4. The intended meta-analysis should find out whether certain (types of) metamodels have proven to be generally better understandable for human users. One of the earliest disputes relevant for this question took place in artificial intelligence by praising the merits of either predicate logic [12], which is usually written as text, or visual representations and diagrams [29]. This debate is excluded here from further investigation as it is based only on (quite suggestive) examples and, thus, differs in type from controlled experiments (see [RQ1] in Section2).

Table 2 lists some experiments examining the understandability of (types of) metamodels. The selection of the studies (deliberately) does not satisfy the requirements postulated in Section 2, as it is intended to point out the obstacles for meta-analysis:

The experiments differ in their independent variables and, thus, in the hypotheses (H_a)[1] investigated. Most independent variables are related to metamodels, but refer to *abstract*[2] syntax ([3]: H_a: metamodels with more constructs easier to understand), *concrete*[2] syntax ([8], [14]; H_a: graphical notation is easier to understand) or a *mixture* of both ([5], [6], [20], [22]). In the mixture case, the understandability of particular metamodels (the listed 'levels' in Table 2) is tested, whereas syntactically pure independent variables characterize types of metamodels. Besides the metamodel, also other factors influencing understandability are investigated, e.g., the complexity of the presented artifacts [14] and the knowledge of the participants [23].

The dependent variables are more homogeneous (correctness, time, perceived ease of use), but the particular measures vary. For example, correctness is quantified by the number of correct answers and by reviews. Additionally, diverse experimental designs have been used. Experimental design, i.e., the way participants are selected and assigned to experimental conditions [26], is discussed in Section 4.2

None of the studies in Table 2 reported effect sizes. [3], [6], [8], [20] and [22] provide at least enough aggregated data to calculate the effect sizes ex post according

[1] H_a denotes the alternative hypothesis, which is given in an aggregated and simplified form.

[2] Abstract (concrete) syntax mean the constructs and their allowed connections (notation).

to Table 1. The effects are small [6], medium [3] or large [6], [8], [20], [22]; see Table 2. But, because of the heterogeneous variables and hypotheses, a methodologically sound meta-analysis cannot be conducted.

Meta-analysis of understandability would be facilitated by some guideline for the planning, conducting and reporting of the underlying experiments. The following groups of guidelines have been proposed:

1. *General guidelines* on experimental research in software engineering (e.g., [4], [19]) with 'best practices' for planning, conducting, evaluating and reporting any kind of experiment. They do not help researchers in selecting variables and experimental designs to investigate understandability.
2. *Guidelines on reporting* the results of experiments e.g., [19], [16]. Though the latter ones have recently been criticized [18], they provide a solid foundation for the prospective availability of data needed to calculate effect sizes.
3. *Guidelines* for experiments in the field of *conceptual modeling*, e.g. 23], [2], [11] or management information systems (MIS) research [15]: These guidelines cover specific aspects of metamodel understandability (e.g., the role of domain knowledge) [23], remain vague sets of hints without well-founded recommendations of variables or experimental designs [2] or aim at classifying existing experimental studies [11]. As a consequence, the classification guideline [11] concentrates on variables that have been used in experiments on metamodel understandability, but neglects potential variables known from cognitive psychology, which is the major field for scientific investigations of understandability. Meta-analytic comparability, however, requires the consideration of all known factors affecting some phenomenon. Experimental design is only discussed in the MIS research framework [15]. Because of focusing on the usage of MIS, 'metamodel' is not considered as an independent variable. Corresponding modifications of the framework have been proposed [6], but remain at the surface. Additionally, the MIS research framework differs in terminology and methodology form empirical software engineering.

To sum it up, owing to heterogeneous experiments and deficient reporting of the experimental results, meta-analysis of metamodel understandability is currently not possible. Appropriate reporting guidelines exist. The next section proposes a framework that is to increase the comparability of experiments on metamodel understandability, which is a prerequisite for meta-analysis.

4 A Framework for Comparable Experiments on Metamodel Understandability

4.1 Affecting Factors

An *experiment* is a scientific investigation in which one or more independent variables (IV) are systematically manipulated to observe their effects on one or more dependent variables (DV) [28]. The outcome of an experiment depends on the *affecting factors* [11]. This term comprises both *independent variables* whose (causal) relationship to the dependent variables is examined and other factors (*extraneous variables, EV*) that confound the causal results [28]. Whether some affecting factor

Table 2. Experiments on the Understandability of Metamodels

Ref.	Independent Variables (Levels)	Tasks (Number)	Dependent Variables	Experim. Design	N (T)	Statistical Procedure	Results	Effect
[5]	Data model (EER, RDM)	Spec (1 case)	CO (review), PEU	2 groups, matched in experience	42 (M)	t-test of means	EER leads to higher correctness; no difference in perceived ease of use	not applicable
[6]	Conceptual data model (EER, KOOM)	Spec (1 case)	CO (review)	2 groups	38 (–)	matched-pairs t-test for means	Mostly no differences in correctness; higher correctness of EER only for some facets	d = 0.04 to d = 2.12
[8]	Graphical query languages	Comp (32), Spec (14)	CO (review)	1 group	27 (U)	χ^2-test on distribution	Graphical queries are easy to comprehend, not easy to specify	$\omega = 0.61$
[14]	Datebase representation (graphical, textual), complexity	Spec (20)	CO (review), ST, PEU	2 x 2 factorial	36 (M)	ANOVA	Graphical representations are faster, lead to higher correctness and higher perceived ease of use.	not applicable
[20]	Conceptual data models (EER, SOM, ORM, OMT)	Spec (2 cases)	CO (review), MT, PEU	4 groups	100 (–)	Duncan test	Increased correctness and faster solutions for EER and OMT	$\eta^2 = 0.14$
[22]	Conceptual models (DSD, ERM, OOM)	Comp (30)	CO (answers), ST	3 groups	121 (M)	ANOVA, correlation analysis	Highest correctness for OOM; faster for OOM, followed by DSD, ERM	$\eta^2 = 0.15$
[3]	Conceptual data models (varying construct number)	Comp (40)	a) CO (answers), b) inverse of time, c) learnability	2 groups	64 (M)	t-test of means difference	Models with more constructs lead to more accurate conceptualization, increase the time to process a schema, are faster to learn	r = 0.41
[23]	Conceptual data models (ER, EER), knowledge	Comp (36)	CO (answers and review)	2 x 2 factorial	81 (U)	paired t-test of means	IS knowledge affects problem solving; domain knowledge is helpful in solving demanding tasks	not applicable

Abbreviations: CO: Correctness, Comp: Comprehension, DSD: Data Structure Diagram, EER: Extended Entity-Relationship Model, (K)OOM: (Kroenkes) Object Oriented Model, MT: Modeling Time, N: Total number of participants, PEU: Perceived ease of use, SOM: Semantic Object Model, Spec: Specification, ST: solution Time, ORM: Object Role Model, OMT: Object Modeling Technique, RDM: Relational Data Model, T: Type of participants constitutes an independent or an extraneous variable is, to some extent, a matter of the researcher's decision (contingent on the research question, the availability of participants, costs etc.). This decision requires knowledge on (at most) all the factors that affect the outcomes of an experiment. For experiments on understandability in computer science, this knowledge is provided by Fig. 1.

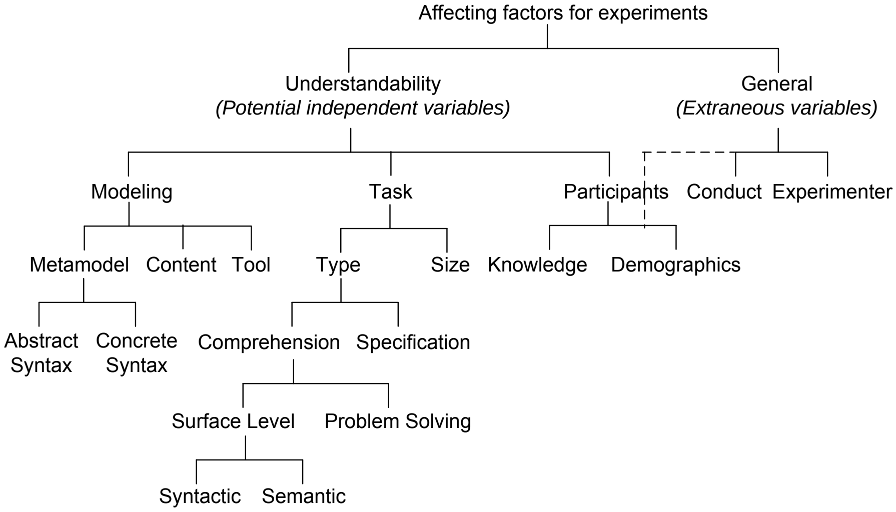

Fig. 1: Affecting factors in experiments on metamodel understandability.

It can be distinguished between factors that affect the outcome of any experiment (*general* affecting factors) and factors with a known influence on understandability; see Fig. 1. In the field of behavioral sciences (to which cognitive psychology belongs), the following *general affecting factors* are acknowledged:

- The *conduct* of the experiment, comprising:
  - The *experimental situation*, namely the location (noise, room temperature), the time of day and the equipment (failures, calibration) [9].
  - *Position effects*: Performance depends on the timely distance of a task from the start of the experiment (e.g., fatigue, getting bored, learning) [21].
  - *Carry-over effect*: The performance achieved in some task depends on whether or not some other task has been done before [28].
- The *experimenter*: His/her ability to instruct participants; his/her *bias* (expecting a particular outcome can distort the experimenter's behavior or data gathering) [26].

These general affecting factors are not causally related to the dependent variables, but distort the experimental results and, thus, are extraneous variables. In contrast, in investigating metamodel understandability, the following affecting factors – related to modeling, participants and task - are potential independent variables (see Fig. 1):

Both the metamodel's *abstract syntax* (e.g., the number [3] or type [7] of constructs) and its *concrete syntax* (graphical vs. textual notation; e.g., [14]) affect understandability. Metamodels cannot be tested in isolation, but only by applying them to some *content.* The content should be 'informationally equivalent' [23], i.e., it must be possible to model this content by any of the investigated metamodels, and the content should be comparably difficult. Finally, the *tool* used to create or present models (e.g., its navigation or dynamic layout capabilities) influences understandability.

Among the affecting factors, *participants* play an intermediate role: Their *demographic* characteristics (e.g., age, gender) affect any experiment [1] and, thus, also understandability. For example, the participants' age is treated as an independent variable in MIS research [15]. *Knowledge* comprises experience and skills related to domain and metamodel as well as general mental abilities. Domain knowledge distorts results on metamodel understandability as it enables inferences [23]. Metamodel knowledge is usually provided in preparing the participants for the experiment.

Tasks in experiments on understandability can be characterized by their type and size. As Table 2 indicates, the task *types* used are comprehension or specification (defined in Section 1), which agree to the dependent variables cognitive psychology suggests (see Section 4.3). Comprehension tasks can be subdivided into surface-level understanding and problem-solving tasks [17]. In *problem-solving tasks,* participants are requested to determine whether and how certain information can be retrieved from an artifact created by applying the metamodel. In contrast, *syntactic* surface level understanding tasks refer to the constructs of the metamodel and their relationships (e.g., 'How many attributes describe the entity type ORDER?'), whereas *semantic* tasks assess the understanding of the *contents* described (e.g., 'Every employee has (a) a unique employee number, (b) more than one employee number.') [17]. An influence of the size of some task, e.g., the complexity of the database described by some metamodel, is generally assumed, but it was only marginally significant in [14].

Depending on the decision of the researcher, a potential independent variable is either systematically manipulated or becomes an extraneous variable. Extraneous variables decrease the *internal validity* of experiments, i.e., the degree to which the variation of the dependent variables can be attributed to the independent variables (rather than to some other factor) [28]. Consequently, extraneous variables must be controlled, which is main constituent of experimental design (see Section 4.2).

4.2 Experimental Design

An *experimental design* can be regarded as a general plan for (types of) experiments that joins independent variables and control techniques for extraneous variables. The main *control techniques* are removing, constancy and randomization [26], [28]; they should be applied in the following order:

1. *Remove* the extraneous variable (EV), especially if it is related to the experimental situation (e.g., use a quite room).
2. If the EV cannot be removed, its influence on the dependent variable is known and the sample is small, keep the EV constant. *Constancy* guarantees that all conditions are identical except for the manipulation of the independent variable, but reduces the *external validity* of the experiment, i.e., its generalizability [26].
3. If sample size does not matter and the influence of some irremovable EV on the dependent variable is not surely known (e.g., gender), must be neutralized (e.g., position or carry-over effects) or should be equated (e.g., age, knowledge), *randomize* the EV. Randomization increases the external validity of experiments.

Table 3. Summary of Experimental Designs

Design	**Between-subjects**	**Within-subjects**	**Block (Matched)**	**Factorial**
No. of IV (levels)	1 (n)	1 (n)	1 (n)	m > 1 (n)
Groups	n	1	n	m × n
Pro:	No carry-over effects	• Simple • Small samples • Constancy of individual characteristics	• Precise • No carry-over effects • Individual differences balanced	Interactions between IV can be examined
Contra:	• Unequal groups possible • Large samples	• Carry-over effects • Experimenter bias	• Effort • Matching factor must exist	• Large samples • Difficult to interpret for m > 3
EV Control	Randomization	Constancy	Constancy and Randomization	Randomization
Statistical test procedures				
Metric DV	♦: independent t ∗: F-test, ANOVA	♦: paired t-test of means ∗: MANOVA		MANOVA
Ordinal DV	♦: Mann-Whitney U ∗: Kruskal-Wallis H	♦: Wilcoxon signed rank test (matched) ∗: Friedman's χ^2		-
Nominal DV	♦/∗: χ^2 contingency test	♦: Sign test, McNemar's test of change ∗: Cochran's Q-test		-
Sample Size ♣	1-t: $n_i = 20$ [50] 2-t: $n_i = 25$ [60]	1-t: N = 11 [23] 2-t: N = 15 [35]	see between-subjects	2-t only, m = 3: $n_i = 20$ [50]

♣ To detect a large [a medium] effect (see Table 1) with $(1-\beta) = 0.8$ and $\alpha = 0.05$.

The *experimental design* to be chosen depends on (1) the number of independent variables and (2) the control technique. Table 3 summarizes typical experimental designs and their (dis-) advantages (for details, see [9], [26], [28]). Experimental design and the dependent variables determine the statistical test procedures for evaluation (see Table 3). For each statistical procedure, an effect size measure exists (see Table 1). The sample size required to detect a small, medium or large effect for a given experimental design and statistical test procedure can be calculated by power analysis (e.g., [9], [10]); the resulting recommendations are given in Table 3.

4.3 Affected Factors

The *dependent variable* is the one on which the effect of the independent variable is measured. Behaviorism, the origin of experimental research in psychology, requires the dependent variable to refer to observable behavior [1]. Thus, 'perceived ease of use' (even though applied, see Table 2) is not an acceptable dependent variable. Instead, the following measures of behavior are common [26]:

1. *Frequency*, e.g., the number of correct answers or solved problems.
2. *Selection*, e.g., which of several answers is chosen.

3. *Response* latency (or response time), which is concerned with how long it takes for a behavior to be emitted, e.g., how quickly a participant reacts.
4. *Response duration*, i.e., the length of time some behavior occurs (e.g., how long a participant deals with a task).
5. *Amplitude*, measuring the strength of response.

The dependent variables in experiments on metamodel understandability (see Table 2) use these measures as follows: Solution time refers to response latency and modeling time to response duration. If correctness is verified by multiple-choice questions (e.g., [17]), it is based on the measure ‘selection’, whereas numbers of correct answers are a measure of frequency.

Thus, the dependent variables in experiments on understandability in computer science are well-grounded in cognitive psychology. Completeness could be achieved by measuring amplitude, which, however, is mainly common in neuroscience [1], and by using selection of some metamodel from a list in specification tasks.

5 Conclusion

Missing comparability of the integrated studies is a major reservation about meta-analysis [27]. But, comparability of heterogeneous experiments can be achieved by methodologically equalizing differences among experiments [13] – provided that the differences are known. In other words, sound meta-analysis is possible if all variables and (for EV) their control techniques are reported. The taxonomies provided by the framework (see Section 4) help researchers to compile such lists; further advances can be achieved by web-publishing them (and the related experimental studies) as well as by tool support for the experiments on understandability. A simple open-source tool called `notate` already exists (http://sourceforge.net/projects/notate). It has been successfully applied in experiments on understandability [24] and can be extended to cover the complete framework of Section 4.

In contrast to the narrow view of MIS research, extensibility and flexibility are major requirements for a framework to investigate understandability in computer science, since the nature of language understanding in general still is an open research question in cognitive psychology [1]. Workshops are an appropriate place to exchange experience in this field and to advance the framework proposed here.

References

1. Anderson, J.R.: Cognitive Psychology and its Implications. 5th ed., Worth, New York (2000)
2. Aranda, J., Ernst, N., Horkoff, J., Easterbrook, S.: A Framework for Empirical Evaluation of Model Comprehensibility. Proc. Intern. Workshop on Modeling in Software Engineering (MISE’07), Minneapolis/ MN. IEEE (2007)
3. Bajaj, A.: The effect of the number of concepts on the readability of schemas: an empirical study with data models. Requirements Engineering 9, 261-270 (2004)
4. Basili, V.R., Selby, R.W., Hutchens, D.H.: Experimentation in Software Engineering. IEEE Transactions on Software Engineering SE-12 7, 733-743 (1986)

5. Batra, D., Hoffer, J.A., Bostrom, R.P.: Comparing Representations with Relational and EER Models. Comm. of the ACM 33, 126–139 (1990)
6. Bock, D., Ryan, T.: Accuracy in Modeling with Extended Entity Relationship and Object Oriented Data Models. J. of Database Management 4, 30-39 (1993)
7. Bodart, F., Patel, A., Sim, M., Weber, R.: Should Optional Properties Be Used in Conceptual Modelling? A Theory and Three Empirical Tests. Information Systems Research 12, 384-405 (2001)
8. Chan, H.C.: Naturalness of Graphical Queries Based on the Entity Relationship Model. J. of Database Management 6, 3–13 (1995)
9. Clark-Carter, D.: Quantitative psychological research. Psychology Press, Hove (2004)
10. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. 2nd ed., Erlbaum, Hillsdale (1988)
11. Gemino, A., Wand, Y.: A framework for empirical evaluation of conceptual modeling techniques. Requirements Engineering 9, 248-260 (2004)
12. Hayes, P.J.: Some Problems and Non-Problems in Representation Theory. Proc. of the AISB Summer Conference. University of Sussex, 63-79 (1974)
13. Hwang, M.I.: The Use of Meta-Analysis in MIS: Research: Promises and Problems. The DATA BASE for Advances in Information Systems 27, 35-48 (1996)
14. Jamison, W., Teng, J.T.C.: Effects of Graphical Versus Textual Representation of Database Structure on Query Performance. J. of Database Management 4, 16–23 (1993)
15. Jenkins, M.A.: MIS Design Variables and Decision Making Performance. UMI Research Press, Ann Arbor (1976)
16. Jedlitschka, A., Pfahl, D.: Reporting Guidelines for Controlled Experiments in Software Engineering. Proc. of ACM/IEEE Intern. Symposium on Software Engineering 2004 (ISESE 2004), 261-270. IEEE (2004)
17. Khatri, V., Vessey, I., Ramesh, V., Clay, P., Park, S.-J.: Understanding Conceptual Schemas: Exploring the Role of Application and IS domain Knowledge. Information Systems Research 17, 81-99 (2006)
18. Kitchenham, B. et al.: Evaluation guidelines for reporting empirical software engineering Studies. Empirical Software Engineering 13, 97-121 (2008)
19. Kitchenham, B.A. et al.: Preliminary Guidelines for Empirical Research in Software Engineering. IEEE Transactions on Software Engineering 28, 721-734 (2002)
20. Lee, H., Choi, B.G.: A Comparative Study of Conceptual Data Modeling Techniques. J. of Database Management 9, 26-35 (1998)
21. Mook, D.: Classic experiments in psychology. Greenwood, Westport (2004)
22. Palvia, P.C., Liao, C., To, P.-L.: The Impact of Conceptual Data Models on End-User Performance. J. of Database Management 3, 4-15 (1992)
23. Parsons, J., Cole, L.: What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modelling techniques. Data & Knowledge Engineering 55, 327-342 (2005)
24. Patig, S.: A Practical Guide to Testing the Understandability of Notations. Proc. 5th Asia-Pacific Conf. on Conceptual Modelling (APCCM 2008). CRPIT Volume 79. ACS, (2008)
25. Pickard, L.M., Kitchenham, B.A., Jones, P.W.: Combining empirical results in software engineering. Information and Software Technology 40, 811-812 (1998)
26. Robinson, P.W.: Fundamentals of Experimental Psychology, 2nd ed., Prentice-Hall, Englewood Cliffs (1981)
27. Rosenthal, R., DiMatteo, M.R.: Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. Annual review of psychology 52, 59-82 (2001)
28. Sarafino, E.P.: Research Methods: Using Processes and Procedures of Science to Underssand Behavior. Pearson, Upper Saddle River (2005)
29. Sloman, A.: Interactions Between Philosophy and Artificial Intelligence. Artificial Intelligence 2, 209–225 (1971)