Science Education
Review Letters

The Terminology within German Lower Secondary Physics Textbooks

Hendrik Härtig[a*]
[a] IPN - Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik , Kiel, Germany

[*] Corresponding author: haertig@ipn.uni-kiel.de

Received 21 October 2013, Accepted 5 May 2014

Abstract
From primary school through university and beyond, textbooks are crucial for teaching and learning science. In most cases the content of the books is aligned with intended state curricula, and teachers in many countries use textbooks at least to guide lesson preparation. Therefore, textbooks are discussed as an implemented curriculum. Many existing studies have been conducted on the content of science textbooks. Using different methods, Biology textbooks as well as Physics textbooks could be shown to be not very coherent in sense of using striking scientific terms regularly. However from the viewpoint of learning only the frequent use of a term will elicit stable propositions around the term, which allows emerging concepts associated with the terms in the mental representation of the readers.

Keywords
Textbooks, terminology, learning

Introduction and theoretical background

Numerous authors state that textbooks are crucial for teaching and learning science. Perhaps not surprisingly teachers in many countries use textbooks at least to prepare their lessons (Kirk, Matthews, & Kurtts, 2001). Especially beginning and prospective teachers believe that they need to follow textbooks and teachers' guides if they want to become successful teachers (Davis & Krajcik, 2005). Teachers use the textbooks to decide what content to teach and how to present it (Schmidt, McKnight, & Raizen, 1997). Furthermore textbooks are used in lessons for activities such as reading excerpts, working on tasks, and so on (Naughten, Schreck, & Heikinnen, 2008; Ball & Cohen, 1996). Textbooks in Germany are commonly aligned with intended state curricula due to the requirement of passing a state accreditation (Valverde et al., 2002). However there are some cultural differences, as curriculum is used quite differently in several countries and contexts. Valverde et al. (2002) differentiate in an international overview between an intended curriculum, which is usually passed by the government and the implemented curriculum which takes place in the classroom.
In many countries it might be true, that textbooks should not only be viewed as intended (and extended) curriculum (Valverde et al., 2002), but rather as implemented curriculum. For Germany this is only partly correct, as teachers do not regularly use the textbook within science class. German physics teachers

typically plan their lessons using textbooks (Haertig, Kauertz, & Fischer, 2012) but usually do not use them during classroom instruction. While there is no study on the usage of physics textbooks, Beerenwinkel and Gräsel (2005) found that the majority of German chemistry teachers use textbooks at a maximum of once per month within their classes - which might also be true for German physics teachers. On the one hand Haertig (2010) could show that the use of terminology in textbooks is somehow comparable to the use of terminology in videotaped physics lessons, but on the other hand, in Germany each student in lower secondary education gets a physics textbook for free (or a small fee) to allow the students preparing physics classes. But this will only happen if the textbooks are written in a readable manner (cf. Haertig & Neumann, 2014). But more than 20 years ago, Merzyn (1987) conducted a study on the terminology of German physics textbooks and found that they overwhelm students with way too many different terms. In the past, this led to the assumption that students will not learn too much about physics from German textbooks, neither during science class nor at home.

As textbooks are seen as (partly) implemented curriculum, much research has internationally been conducted analyzing their content. Some former studies focus on the terminology; and even if some authors speak of concepts, they code terms. Shavelson (1971) utilized directed graphs, which are a technique by which the content structure of texts can be represented. He made a list based on the frequency of concept usage within one text on mechanics, and analysed the presented associations in the text with the digraphs. Shavelson found that the number of learned correct associations according to a specific concept is highly correlated with the frequency of this concept's occurrence in scientifc texts. These results have been reproduced for other content areas of physics (Geeslin & Shavelson, 1975) as well as for mathematics (Thro, 1978). More recently Nehm et al. (2008) used terms to examine the usage of specific important concepts in Biology textbooks. They concluded that despite being highly relevant from an expert point of view, these terms are not used frequently within textbooks. Rather than being used regularly to connect different chapters or units, concepts were segregated within particular chapters.

A commonality of the above studies is a focus on connections between content structure presented in textbooks and learning. The studies provide evidence about the importance of content structure in texts due to the influence of content structure on students' learning. While these studies all focused on very small coding units - single words in most cases - the majority of these content analyses were related to concepts. This connection is based on theories about learning from written texts. Johnson (1964) points out that the meaning of a concept arises through the associations of, or relations to, other concepts. By reading associations and learning associations, cognitive structures between different terms are forged (Thro, 1978). Through this process, terms grow into concepts. Johnson (1971) concluded that learning, at least in physics, includes building associations between important terms. This might be another reason why an exhaustive use of different terms may hinder learning: Too many different terms in one (part of a) textbook cannot be connected.

With respect to these cognitive assumptions we follow former works and analyze the terminology of physics textbooks (cf. Shavelson, 1972; Merzyn, 1987; Nehm, 2008). According to Nehm (2008) we prefer to use terms as entity of coding instead of concepts aligning with learning theories (e.g. Deese, 1962; Ausubel, 1963; Gagné, 1985) in that terms are the smallest entity that a learner would be confronted with. Former studies offer only partly evidence concerning the terminology of physics textbooks: First, the only comparable study from Merzyn (1987) analyzed the books by counting different terms more than twenty years ago. Second, many other studies focus on some specific terms / concepts which have been chosen by experts in advance. Therefore this study is aimed at replicate Merzyns results and enhancing approaches like that from Nehm, who analyzed the occurrence of terms throughout different chapters. By analyzing all terms and quantifying them one will be able to not define in advance which might be the most relevant ones but to analyze this aspect afterwards.

Research Questions

As textbooks are of high importance for teaching and learning science, it is critical to analyze how the content is presented to the students. Much research focuses on terms or concepts as these are seen as smallest entity when students learn using texts. So far existing studies are more than twenty years old or concentrate on the occurrence of few terms, which are defined to be central in advance. This study will bring both approaches together: An analysis technique for physics textbooks shall take all terms into account. This should first prove the replicability of the overwhelming amount of terms in former studies and second try to lead to a empirical base for deciding which terms are indeed of high importance. The two guiding research questions are:

1) Which measures are appropriate getting to know how regularly specific terms are used within physics textbooks?

2) Which terms are - based on the measures - used very regularly and might therefore be seen as important?

Method

In Germany, science is mostly taught as separate subjects: physics, chemistry, and biology from grade 7 to grade 10. This study concentrates solely upon physics. Due to the large number of physics textbooks used in Germany, our sample consists of four different textbooks. This selection was based upon the common procedure of each publisher having a master edition of their unique textbook, so that individual chapters can be added or removed to comply with differences in the curriculum of German states. One textbook was selected from each of the four biggest publishers in Germany (Bredthauer et al., 2002; Kuhn (Ed.), 2003; Schön & Wilcke (Eds.), 2002 and Mikelskis, Schön, & Wilcke (Eds.), 2001 as one book in two parts; Boysen et al., 1991). All of these textbooks covered grades seven to ten. This sample has been shown to be representative for other German physics textbooks analyzing the contained units in the four chosen textbooks and comparing them to other randomly selected books.

As entity of coding, a term is defined as a "verbal designation of a general concept in a specific subject field." (ISO 1087-1 (2000:6)). Based on this definition, nouns in a given textbook are 'terms' if they are clearly defined within the subject of a textbook or a chapter. Here an important linguistic aspect should be noted: In German, terms are really single words while for example in English that is not true (e.g. "steam engine" is "Dampfmaschine"). Therefore in English one would have to slightly adopt the coding procedures.

With respect to the research questions and former analyses of textbooks terminology three values are considered to represent how regularly a specific term is used: (1) the total number of term occurrence, (2) the distance between the first and the last usage of a specific term, and (3) the standard deviation of pages between two term usages. The first one is directly adopted from the work of Shavelson (1972) or Merzyn (1987) who just count all terms. The second and the third one are related to the approach of e.g. Nehm (2008) or Chiappetta, Fillman, and Sethna (1991a; 1991b). They want to find out, whether specific terms or concepts are used either at all or regularly throughout specific topics. As German textbooks are written in a very hierarchical way and clearly split by topics (e.g. mechanics from page 10 to 80 and electromagnetism from page 81 to 145) the information of total distance and standard deviation of distances between usage let us measure how regularly a term is used throughout the book: Image one term is used 20 times throughout the book, but 10 times in the first chapter (e.g. "light" in optics) and then again 10 times in the last chapter (e.g. astronomy) the distance between first and last usage is really high, however the standard deviation will be high, which also tells us this term is not used regularly within all topics but in two which are far away from each other within this book.

Procedures

To ensure a reliable procedure, the definition for a term has been operationalised and illustrated within a coding guideline. Two steps have been separated: (1) Deciding which words are possible physics terms and which are not; (2) coding the textbooks themselves. For the first step the indexes of several physics textbooks have been scanned and saved as a long list that contains all entries. Using the first part of the coding guideline (see supplemental material online), four experts rated each term as to whether an entry might be a physics-term (in a physics context) or not. All experts held at least a master's degree in learning and teaching physics, and some were PhDs. The term definition was operationalised and illustrated by using a list of terms (cf. Table 1). The experts agreed on this operational definition:

A noun (perhaps restricted by an adjective) or its pronoun is a scientific term if it is well defined in physics and used in a physics context. (Again, it is important to note that in other languages than German more than one noun can make-up one term. However in German this is not the case.) The decision is made based on university-level physics. Two limitations of this approach are noted: Measurement equipment and experiments are only coded as physics

Table 1: *Exemplary rating of index entries from a physics textbook (Hewitt, 2010).*

Index entry	*Term category*
Icebergs	No term: everyday object
Ideal efficiency	Term: variable / constant
Illusions, optical	No term: everyday object
Image, optical	Term: Principle / Procedure
Image, virtual	Term: Principle / Procedure

Note. This table serves as example for the database. Within an expert rating for each index entry it has been decided, whether it is a term or not. For both different subcategories are used to reach a reliable procedure. For example measurable and well defined variables such as voltage or distance are coded in the category *Term: Variable / Constant*. Words which are much more relevant in everyday life or in other subjects are coded to be not a term like "ice". If you would like to find reasons for the specific categories please take a look at the coding guide which is provided as supplementary material.

terms if they are used to measure explicitly one physics variable or if they illustrate a relationship between two variables. Additionally, technical equipment is not included as a scientific term if it is primarily relevant to a student's daily life.

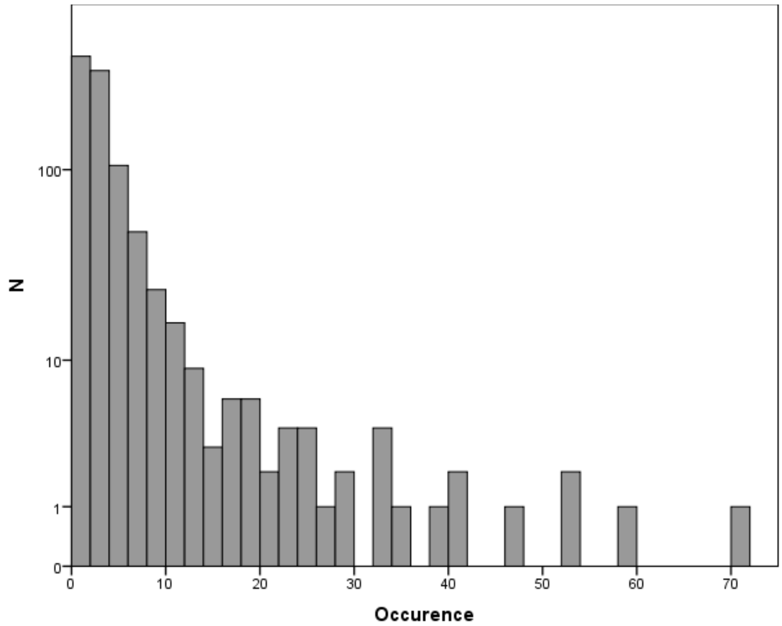
The second step was the coding itself. Students with a bachelor in physics were trained using the guideline, parsing physics terms, and coding the number of the page where a specific term is mentioned.

Results

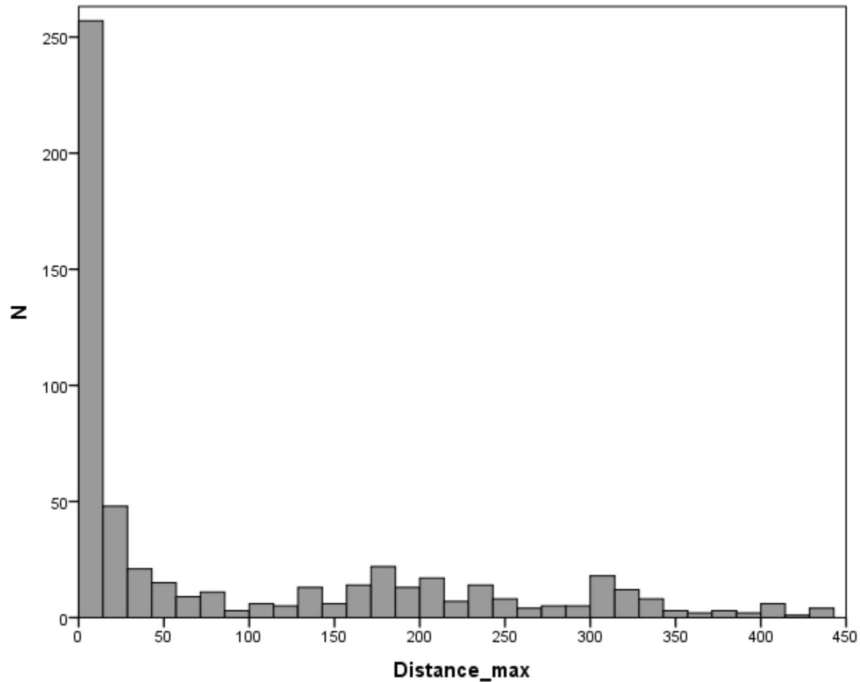
First, the reliability of separating terms was proven. While one expert rated all index entries according to our coding scheme, other experts rated independently a random sample of about 10% of all index entries. The analysis of interrater agreement between the experts indicates that our construct delineation was supported (Fleiss' kappa: κ=.588; p<.001; pair-wise comparison: Cohen's kappa κ=.64<κ<.748; p< .001 (N=265)). This interrater agreement is on a mediocre level and a reasonable benchmark for analyzing textbooks does not exist. Reviewing former studies on the analysis of textbooks' content, only few report interrater agreements as measurement for the quality. Most of those use the percentage agreement (Chiang-Soong & Yager, 1993; Stern & Roseman, 2004; Roseman et al., 2010) and report values between 80% and 97%. However the percentage agreement is not corrected for agreement by chance and is therefore not reliable. Only Chiappetta, Fillman, and Sethna (1991a; 1991b) also report a pair-wise calculated kappa which exceeds .700 and is thereby comparable to our results.

As the interrater agreement is not perfect and the coding of the textbooks shall not be influenced by different expertise, after the rating from the main rater all physics terms and non-physics terms out of the four textbook indices were entered into the database including the category they shall be coded for (see table 1). Based on this database all four textbooks have been coded. Additionally if one term which has not been part of the indices occurred while coding the books, the database has been extended and the other textbooks have been scanned for this term, too. After coding four textbooks, the list of terms contained 4151 entries, including 2172 (52.3%) physics terms. The four textbooks contain 335 to 367 pages and 704 to 1067 terms. The number of new terms per page ranges between 1.92 and 2.95. The correlation between the term occurrence was above Pearsons r =.900; p <.001 between the four textbooks, they are highly similar. For this reason here will be presented only the results for one exemplary textbook and the generalisability will be discussed.

The exemplary chosen textbook contains 939 different terms. However their absolute number of usage differs a lot. 377 of these terms are used only once within the textbook. The mean of term usage is 3.73 and the most used term ("Koerper" which means "physical body") is used 70 times throughout the whole book. As the standard deviation of term usage is 6.186, only 66 terms are used more often than outside the first standard interval (about 10 usages). This is the reason, why the graph for term occurrence has to be plotted using a logarithmic scale for the y-axis: The majority of terms are used really seldom which perfectly fits former results from Merzyn (1987) (cf. graph 1).



Graph 1: Number of term occurence in physics textbook. As the y-axis (number of terms with a specific number of occurrences) uses a logarithmic scale, one can see that the vast majority of terms is used really seldomly.



Graph 2: Maximum distance of a term in physics textbook. The majority of terms have a maximum distance below 50 pages, which means that they will be typically used only within one topic.

Concerning the coverage of terms throughout the textbooks one recognizes again that the majority of terms is used in a narrow area - in graph two those terms which occur only once will not appear, however there is still a huge amount of terms which cover only about 50 pages (mean is 56.5 pages). This time 168 terms have a span of usage within the book that is higher than the first standard deviation interval. Taking a closer look at the standard deviation of pages between two usages for a specific term additionally gives us the supposed details: some of the terms that cover a high absolute distance are used not very regularly. For example “sun” indeed is used very early and very late within the book, which might be considered being good, but since standard deviation is pretty high, this does not represent a frequent usage of the term that is useful for building a concept over time. For example the term "model" reaches similar values for absolute occurrence and maximum of distance but only half the value for the standard deviation, which might be better, as it is used pretty regularly.

With respect to the second research question it might be interesting, which terms are those with very values for absolute occurrence, maximum distance and low value for standard deviation of distance between two occurrences. Therefore we searched for those terms with a higher absolute occurrence than medium plus one standard deviation, higher maximum distance than medium plus one standard deviation and lower standard deviation between two occurrences than medium plus standard deviation. Only 39 term out of the 939 fulfill these criteria. These might reasonably be seen as the most important terms of the textbook as they are used really often regularly throughout the whole book. The top ten is: body, current (as the phenomenon: flow of charges), voltage, energy, electron, force, direction, light, second, and meter.

Discussion

As students' learning from texts is highly dependent on the presented content structure (Shavelson, 1971), it is not surprising that the last decade has witnessed many new studies on textbook content (e.g. Roseman et al., 2010). If the content structure of the textbooks should facilitate learning, there would be a reasonable number of terms which serve as real concepts by anchoring very different topics through being used very often. However our results show that there are only a few highly connected terms in the physics textbooks (typically about 10% of all terms in each textbook). Most terms that are not used very often differ significantly between the four analyzed textbooks. These results indicate a type of core knowledge which consists of those 10% of terms and their interrelations. All the other terms seem to consist of different synonyms, contexts, examples, and so on. However it should be explicitly stated that compared to the study of Nehm et al. (2008) some terms were found to be highly linked and thereby offering coherence. And it is a striking feature, that the overarching ideas from the German nationwide standards can somehow be seen as represented: energy is one of the 20 best linked terms, force can be seen as representation of interactions. Even if here within we presented results only from one textbook, the analyses for three other textbooks are highly comparable, as the three measures are highly correlated between the four textbooks.

Based on our results it might be argued that the textbooks indeed might promote the learning of overarching ideas in some cases. Nevertheless, the vast use of different terms and synonyms might hinder the students in gathering the important elements to build up a mental representation. The textbooks' content covers huge areas, albeit weakly connected. Hence, many terms are used very rarely, and the importance of teaching each topic in detail should be debated. It seems as if a broader view and more interconnections between the different topics would be possible with fewer terms. Fortus and Krajcik (2012) call this language coherence: "In coherent curricula it is important either to use terms in a consistent manner across all contexts or to explicitly clarify the different meanings the terms have in different places. Why they are used in one place in one way and a different way in another place." However there is one important limitation of this study - following the works of e.g. Merzyn (1987), or Shavelson (1972), only terms are counted without thinking about the relationships that are made up between the terms. Therefore we are not able to hypothesize about the overall coherence. The analyses presented in this paper focus specifically on terminology, but language is where textbook learning starts.

References

Ausubel, D. P. (1963). Cognitive structure and the facilitation of meaningful verbal learning. In: Journal of Teacher Education, 14 (2), 217-222.

Ball, D. L. & Cohen, D. K. (1996). Reform by the Book: What Is - or Might Be - the Role of Curriculum Materials in Teacher Learning and Instructional Reform. Educational Researcher, 25 (9), 6-8, 14.

Beerenwinkel, A. & Gräsel, C. (2005). Texte im Chemieunterricht: Ergebnisse einer Befragung von Lehrkräften [Texts in chemistry classes: Results of a teacher survey]. Zeitschrift für Didaktik der Naturwissenschaften, 21-39.

Boysen, G., Glunde, H., Heise, H., Muckenfuß, H., Schepers, H. & Wiesmann, H. (1991). Physik für Gymnasien – Sekundarstufe 1. Berlin: Cornelsen.

Bredthauer, W., Bruns, K.G., Müller, W., Klar, G., Schmidt, M., & Wessels, P. (2002). Impulse Physik – Mittelstufe. Stuttgart: Klett.

Chiappetta, E. L., Sethna, G. H., & Fillman, D. A. (1991). A Quantitative Analysis of High School Chemistry Textbooks for Scientific Literacy Themes and Expository Learning Aids. Journal of Research in Science Teaching, 28 (10), 939-951.

Davis, E. A. & Krajcik, J. S. (2005). Designing Educative Curriculum Materials to Promote Teacher Learning. Educational Researcher, 34 (3), 3-14.

Deese, J. (1962). On the structure of associative meaning. Psychological Review, 69 (3), 161-175.

Gagné, R. (1985). The Conditions of Learning (4th ed.). New York: Holt, Rinehart & Winston .

Geeslin, W. E. & Shavelson, R. E. (1975). Comparison of content structure and cognitive structure in high school students' learning probability. In: Journal for Research in Mathematics Education, 6 (2), 109-120.

Haertig, H. (2010). Sachstrukturen von Physikschulbüchern als Grundlage zur Bestimmung der Inhaltsvalidität eines Tests. [Content of Physics Textbooks as Indicator for Test Validity]. Berlin: Logos.

Härtig, H., Kauertz, A. und Fischer, H. E. (2012). Nutzung von Schulbüchern zur Unterrichtsvorbereitung in Physik. [Usage of science textbooks within preparation of science classes]. MNU – Der mathematische und naturwissenschaftliche Unterricht, 65 (4), 197-200.

Haertig, H. & Neumann, K. (2014). Mit Schulbüchern lernen? Wie Schulbücher den Wissenserwerb fördern können [Learning with physics textbooks - how textbooks enable physics learning]. Praxis der Naturwissenschaften Physik in der Schule.

Hewitt, P. G. (2010). Conceptual Physics (11th Edition as international edition). Glenview, IL: Pearson.

ISO 1087-1:2000. International Organization for Standardization. Retrieved 2012, from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=20057.

Johnson, P. E. (1964). Associative Meaning of concepts in physics. In: Journal of Educational Psychology, 55 (2), 84-88.

Johnson, P. E. (1971). Perceptual Structure of Scientific Knowledge. Final Report Project No. 0-0067. U.S. Department of Health, Education, and Welfare.

Kirk, M., Mathews, C. E., & Kurtts, S. (2001). The Trouble with Textbooks. The Science teacher, 68 (12), 42-45.

Kuhn, W. (Ed.) (2003). Physik 1. Braunschweig: westermann.

Merzyn, G. (1987). The language of school science. International Journal of Science Education, 9(4), 483-489.

Mikelskis, H. F., Schön, L.-H., & Wilke, H.-J. (Eds.) (2001). Physik Plus – Gymnasium Klasse 8. Berlin: Volk und Wissen.

Naughton, W., Schreck, J., & Heikkinen, H. (2008). Seeking Evidence for "Curricular Relevancy" within Undergraduate, Liberal Arts Chemistry Textbooks. Journal of Research in Science Teaching, 45 (2), 174-196.

Nehm, R., Poole, T. M., Lyford, M. E., Hoskins, S. G., Carruth, L., Ewers, B. E., & Colberg, P. J. S (2008). Does the Segregation of Evolution in Biology Textbooks and Intro-ductory Courses Reinforce Students' Faulty Mental Models of Biology and Evolution? Evolution: Education and Outreach online. Retrieved May 13, 2010 from http://www.springerlink.com/content/vnt5075254900957/.

Roseman, J. E., Kesidou, S., Stern, L., & Caldwell, A. (1999). Heavy Books light on Learning. Science Books & Films, 35 (6), 9-13.

Roseman, J. E., Kulm, G., & Shuttleworth, S. (2001). Putting Textbooks to the Test. ENC Focus, 8 (3), 56-59.

Roseman, J. E., Stern, L., & Koppal, M. (2010). A Method for Analysing the Coherence of High School Biology Textbooks. In: Journal of Research in Science Teaching, 47 (1), 47-70.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). A Splintered Vision. An Investigation of U.S. Science and Mathematics Education. Dordrecht, Boston, London: Kluwer Academic Publishers.

Schön, L.-H. & Wilke, H.-J. (Eds.) (2002). Physik plus, Gymnasium Klassen 9 / 10. Berlin, Volk und Wissen.

Shavelson, R. J. (1971). Some Aspects of the Relationship between Content Structure and Cognitive Structure in Physics Education. Ann Arbor, U.S.A.: University Microfilms.

Thro, M. P. (1978). Relationships Between Associative and Content Structure of Physics Concepts. In: Journal of Educational Psychology, 70 (6), 971-978.

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). According to the Book. Using TIMSS to investigate the translation of policy into practice through the world of textbooks. Dordrecht, Boston, Lon