WORKING ON GRAMMAR AT SCHOOL: EMPIRICAL RESEARCH FROM GERMAN-SPEAKING REGIONS

REINOLD FUNKE

Institut für deutsche Sprache und Literatur, Pädagogische Hochschule Heidelberg

Abstract

The article reviews quantitative as well as qualitative research on grammatical learning at school in the context of L1 education in German-speaking regions. One of the specifics of this research is that its focus has been not on the effects of grammar learning on reading and writing outcomes but instead on the quality of the knowledge which is gained. Research results are reported with respect to five areas: (1) interventions aimed at promoting grammar knowledge; (2) interventions aimed at fostering linguistic skills; (3) observational studies; (4) large-scale studies; and (5) classroom discourse in grammar lessons. In its conclusion, the article points to issues pertinent to future research. They include measuring the extent to which students access syntactic features directly, without having them explain these features verbally, as well as conceiving of instructional methods to make sure that students gain continued access to syntactic information.

Keywords: grammar instruction, grammatical knowledge, metalinguistic achievement, German.

Funke, R. (2018). Working on grammar at school: empirical research from German-speaking regions. Contribution to a special issue Working on Grammar at School in L1-Education: Empirical Research across Linguistic Regions. L1-Educational Studies in Language and Literature, 18, p. 1-39. https://doi.org/10.17239/L1ESLL-2018.18.04.05

Corresponding author: Reinold Funke, Institut für deutsche Sprache und Literatur, Pädagogische Hochschule Heidelberg, Im Neuenheimer Feld 561, 69120 Heidelberg, Germany; email: funke@ph-heidelberg.de

1. INTRODUCTION

The topic of the present review is grammatical learning contingent upon L1 instruction, its scope being restricted to L1 education in German-speaking regions. In German-speaking regions, L1 education in school is delivered within the context of a school subject called German studies (*Deutschunterricht*). This subject is established in Germany, Austria, the Germanophone parts of Switzerland and some European regions with German-speaking minorities, such as South Tyrol (Italy) and Transylvania (Romania). It gradually replaced Latin studies and Catechism studies as the leading school subjects in higher and in lower education, respectively, during the 19th century. In conformity with the general continental European tradition (Delnoy, Herrlitz, Kroon & Sturm, 1988; Delnoy & Kroon, 1990), grammar instruction has been a part of German studies from the very beginning, despite massive criticism (Grimm, 1819; Ingendahl, 1997; Weisgerber, 1950). Nowadays, according to the educational standards established in German-speaking countries (BIFIE, 2006a, 2006b; EDK, 2007; KMK, 2004, 2005), grammatical competences still form part of the curriculum. They figure under titles such as 'Analyzing Language and its Use' (*Sprache und Sprachgebrauch untersuchen;* KMK, 2004) or 'Considering Language' (*Sprachbetrachtung*; BIFIE, 2006a).

At the heart of the traditional grammar curriculum are syntactic categories (*Wortarten*)—which students need to be familiar with to some extent in order to learn spelling—and sentential constituents (*Satzglieder*) and their functions (i.e., figuring in a sentence as a subject, an object, an adverbial, or a predicative). Other grammatical topics generally listed in the educational standards include sentence patterns and types of complex sentences as well as the inflection of verbs, nouns and adjectives.

According to specifications which Symann (1980) found in the syllabi of the former German Democratic Republic, in the 1970s about 15% of the instructional time in German studies was dedicated to grammar. In a recent survey, primary school teachers in Germany reported that they devoted 16-17% of instructional time to the domain of analyzing language and language use (Helmke & Hosenfeld, 2007). This suggests that the time spent on considering language in German studies has remained constant although the contents of instruction may have changed.

Teaching grammar in what is called 'L1 education' serves not only students for whom the language of schooling is their first language (L1), but also students for whom it actually is their second language (L2). In both cases its primary aim is not to build grammatical competence but rather to increase grammatical knowledge. Grammatical competence embraces the command of syntactic constructions; it is assumed to be given in someone who felicitously makes use of these constructions when communicating or understanding others' communications. By contrast, grammatical knowledge of syntactic constructions is assumed to be present in someone who recognizes such constructions when they occur. Acquisition of

grammatical knowledge does not automatically come along with the acquisition of grammatical competence. In this sense, it is a metalinguistic achievement.

Germanophone research on grammar learning in L1 education has been driven mainly by the question of what active and efficient grammatical knowledge is. In other words, it has investigated the conditions under which students, as speakers and listeners, reliably recognize syntactic constructions by responding to them when they occur. They obviously do so when working on specific assignments (e.g., in deciding whether to write *the student's* or the *students'* in a text); they also do so, albeit less obviously, when they attentively interpret syntactic constructions (e.g., by anticipating how a formulation will continue). They may be said to reliably recognize a syntactic construction if they are able to respond to it in the same way across various instances of it. Indeed, in what follows, evidence of consistent response to syntactic constructions will be considered the criterion for the presence of active and efficient grammatical knowledge.

Germanophone research's orientation towards the conceptual question of what active and efficient grammatical knowledge is diverges to some degree from international research, most notably from Anglophone research, which has focused on the question of whether teaching grammar fosters linguistic skills. Though Germanophone and Anglophone perspectives differ, they may eventually turn out to complement each other. The present review examines the Germanophone research in order to determine the scope and the quality of the grammatical knowledge students gain in L1 education. It then connects its findings to the questions driving Anglophone and other international research, such as whether grammar instruction impacts on writing and whether one should teach grammar at all.

2. METHOD

Asking what grammatical knowledge students acquire in L1 education means raising a factual question and considering a conceptual question along with it: How can one adequately describe the knowledge students are said by the studies to have gained? And does the evidence imply that students gained active and efficient grammatical knowledge according to the criterion set above? In asking such questions, the present review evaluates empirical studies with respect to factual evidence, while also considering how the facts may be appropriately described. Consequently, although studies will be reported according to the questions their authors raised, their findings will, where appropriate, be reconsidered 'from outside', that is, based on the conceptual question. As a result, although this review proceeds through studies, the basic unit of analysis is not the study but the finding.

The review focuses on research conducted on working on grammar within the school context. It excludes developmental studies done outside schools (e.g., those analyzing how children spontaneously broach the issue of language in everyday life). Laboratory studies have also been left out (e.g., those measuring how orthographic cues to syntactic structure impact on reading). Demarcation problems

sometimes occurred where instructional studies referred to orthography or to morphology. In these cases, the study was included if the main focus of instruction seemed to be on syntax; otherwise it was not included. Publications considered here span the period from 1965 to 2015. The utility of going back 50 years may seem doubtful at first glance; however, just because a study is old does not mean that its data are invalid; in fact, the data of some of the older cases included herein have never before been given serious consideration, thus warranting their inclusion here.

Starting from these specifications, quantitative as well as qualitative empirical studies were compiled. In order to find studies, first reviews were consulted (Abraham, 2010; Funke, 2014; Kochan, 1975). Second, in an attempt to locate unpublished research, electronic databases were searched using the terms *+Grammatik* (grammar), *+Grammatikunterricht* (grammar instruction), *+grammatisches Wissen* (grammar knowledge) in isolation or combined with *-Fremdsprache* (foreign language), *+Unterricht* (instruction) and/or *+empirisch* (empirical) (http://www.ams-forschungsnetzwerk.at/, http://www.fachportal-paedagogik.de, http://www.dnb.de, http://edudoc.ch, http://www.forschungsportal.ch, http://www.skbf-csre.ch, http://www.gesis.org). Printed bibliographies were used to identify older research (*Gesamtverzeichnis deutscher Hochschulschriften*, *Deutsche Nationalbibliographie*). Some further information on unpublished dissertations in the former German Democratic Republic was found in Friedrich (1996). However, given that there is no systematic index of research literature in the field of German studies, it is possible that some relevant studies were missed.

It was decided not to exclude methodologically weak studies from the outset but rather to include all studies first and then to sort them according to methodological criteria in a second step. This follows from the principle that the finding, not the study, is the unit of analysis. Studies which are methodologically weak may nevertheless contain valid findings, i.e. findings which are well documented and for which the conditions under which they emerged are described with sufficient precision. The inclusion of such findings is vital for educational research because a methodologically weak study, if rooted in authentic instructional experience, may contain educationally instructive results.

3. RESULTS

3.1 Intervention studies aimed at promoting grammatical knowledge

Overview. Studies under this heading comprise quasi-experimental studies which set out to compare the outcomes of diverse forms of grammar instruction. Most of them are dissertations originating in the former German Democratic Republic during the period from 1968 to 1988. As a rule, their research methodology does not correspond to current standards. Their methodological weaknesses may be summarized as follows: They do not make use of the fact that their data are

longitudinal (applying inferential statistics, if any, on the results of pretests and posttests in isolation); in some cases, they test the same hypothesis repeatedly by using dependent data; and they employ self-conceived pre- and posttests without checking the quality of the tasks. Beyond that, it may go without saying that the instructional objectives inherent in the older studies are to some extent outdated because they rely on linguistic varieties and grammatical tenets which have meanwhile become obsolete. Studies will be reported in three steps: work based on activity theory (Herrmann, 1974; Jantos, 1971; Schößler, 1981; Schumer, 1988), studies rooted in linguistics (Altendorf, 1969; Friedrich, 1970; Meyer, 1969; Ploog, 1981; Siegert, 1977), and residual studies not related to one another (Döhmann, 1977; Kramarczyk & Wazel, 1968; Metz, 2005).

Studies based on activity theory. These studies share the idea that grammatical knowledge is ready for use so long as it is conceptual. According to the Soviet psychologist Galperin (1972, first published 1957), conceptual learning starts with manifest actions ('unfolded materialized action'), and progresses through symbolic actions controlled by linguistic clues to internalized mental actions. Children take this path when they learn to regulate their actions. In order to support this, the studies required teachers to supply children with an 'orientation base', which in practice turned out to be a sheet which listed the steps children had to follow when dealing with a task. Instruction based on this understanding of activity theory is a highly directed process.

Jantos (1971, 1975), Herrmann (1974) and Schumer (1988) implemented this instructional approach in grades 1-2. Jantos's (1971) experimental instruction extended over the whole of the second grade. It focused mainly on the segmentation and classification of verbal and nominal inflection morphemes. An age-matched control group was instructed according to the then-current curricular guidelines. In his final tests, Jantos found a sustained superiority of the experimental group in tasks on inflectional morphology (effect sizes for various tasks scatter around d = .60). The experimental group also outperformed the control group on tasks requiring the identification of verbs (d = .37). With respect to the identification of sentential constituents, no advantage for the experimental group was found when it was compared with a control group consisting of grade 4 students whom Jantos enlisted because sentential constituents had not been a topic of instruction in the age-matched control group (d = -.28). Jantos explains the absence of success in tasks on sentential constituents by the fact that the experimental instruction led children to attend to features of single words instead of attending to syntactic relations. From data he reports about one of his assignments one may indeed estimate that the children in his experimental group identified the subjects of sentences in 90.6% out of all cases and the predicates (which, in the sentences given, consisted of finite verbs) in 68.6%. However, in just 62.4% out of all cases do they seem to have simultaneously identified a sentence's subject and the finite verb. This amounts to the value one would expect if they considered the finite verbs independently of the subject. From a syntactic point of view, however,

to identify the one would necessarily imply the identification of the other because subjects are subjects by virtue of being related to a finite verb via agreement.

Starting as early as in grade 1 and terminating in grade 2, Herrmann (1974) taught two experimental classes on sentence types, verbs and nouns, sentential constituents and patterns of word order. The posttest results favored the experimental classes as compared to control classes (which, depending on the topic, were recruited from grades 1-5) in all domains listed above. However, Herrmann's study did not include a pretest.

Schumer (1988), in instructional experiments dealing with adjectives in grades 2 and 3, focused on the development of gnoseologic interest (*Erkenntnisinteresse*) in grammar. At the end of her trials, the students in the experimental groups showed a more marked interest in grammar than the controls. They were also superior in identifying adjectives and finding out which expressions they modify; moreover, the degree of superiority was related to the quality of the gnoseologic interest they displayed as determined by the reasons they gave for it. However, the effect was restricted to tasks drawing on metagrammatical knowledge. In tasks focusing on the proper use of adjectives, no effect was found. Pretests were lacking in Schumer's experiments as well.

With respect to more advanced students, Galperin's activity theory was made the basis of an instructional experiment by Schößler (1981) in grade 4. Schößler's finding was that the experimental group outperformed the control group by far in the recognition of sentence constituents and their functions. This, however, seems dubious when one considers the test results: Whereas in the pretest the median of correct solutions across all tasks was 80% in both the experimental and the control group, in the posttest it was only 14% in the control group but 84% in the experimental group. This suggests that the main effect consisted of the control group worsening sharply and thus raises suspicion that the posttest was somehow unfair, possibly because the control students had not been taught its terminology.

Studies rooted in linguistics. The formation of conceptual grammatical knowledge was the main objective of a second cohort of studies which stand out by being based more in linguistics than in psychology. The core task of grammar instruction was understood by them as providing students with operational identification procedures (*Identifikationsverfahren*) enabling them to decide whether a grammatical feature is present in a given linguistic unit or not. In addition, most of these studies shared the pedagogical ambition of facilitating progress in weak learners in particular.

In an intervention trial on sentence constituents, Meyer (1969) implemented the linguistic approach in grade 5. He reports that although the experimental students initially struggled with it, in the posttest they were significantly superior to the control group in tasks requiring them to segment and classify sentence constituents. No significant superiority was found in tasks whose solution did not depend on the use of the linguistic procedures taught, such as determining verbal tense.

In two experimental trials in grade 5, the first of which dealt with the recognition of past participles and the second with the recognition of infinitival groups, Siegert (1977) compared more formalized and less formalized versions of identification procedures. In the posttests, no differences were found between the more and less formalized versions. However, when compared to a control group which received instruction according to the common curriculum, both experimental groups performed better. This pertains to hits (that is, classifying participles as participles and infinitival groups as infinitival groups) and, in the case of participles, to false alarms (that is, classifying non-participles as participles). The instructional method applied by Siegert is reminiscent of the 'mastery learning' approach which was popular in the 1970s (Bloom, 1968). In this approach, exercises on a topic are continued until a pre-established ratio of correct solutions is reached by all students. In a final interview it became evident that weak learners felt comfortable with this whereas strong learners had the feeling of being held back by the repeated stepwise execution of the identification procedures which Siegert required from them.

Participial and infinitival groups were also the topics of intervention studies done by Altendorf (1969) and Friedrich (1970) in grade 5. In the experimental condition, both concepts were deliberatively contrasted, whereas in the control condition they were taught in isolation. The effects of contrasting versus isolated instruction were additionally explored in experiments dedicated to the identification of causal and conditional adverbials. Contrasting instruction performed better than isolated instruction in all experiments with respect to false alarms. With respect to hits, either no difference (Altendorf) or a less marked difference (Friedrich) was found. Altendorf and Friedrich followed a 'mastery learning'-like approach as well.

Ploog (1981) conducted an instructional experiment in grade 5 in which students were taught rather formal procedures, which she called 'algorithms', to identify the finite verbs within verbal chains, verbal tenses and verbal voices. Besides this, the experimental instruction also featured a high proportion of individualized work. The control group received 'traditional' instruction. The posttest revealed advantages for the intervention group in all domains of the experimental instruction. Ploog also compared the instructional effects in various achievement groups and found that underperforming students profited more than efficient learners; however, she did not validate this with inferential statistics.

Residual studies. Kramarczik & Wazel (1968) tracked the students of two classes longitudinally from grade 4 to grade 6 to investigate the impact of programmed instruction (in print) about verbal inflection and sentence constituents. In the topics of instruction, the experimental classes were superior to control classes (which varied from year to year). The authors found a rank correlation of .63 between the test scores of the experimental students at the end of grade 4 and at the end of grade 6 (N = 48).

In a study done in West Germany with adults participating in occupational retraining, Döhmann (1977) compared 'traditional' grammar instruction to instruc-

tion based on generative grammar. Döhmann assumed that the latter directly mirrors speakers' mental grammar. In the pre- and posttest, subjects had to assign words to word classes, to segment sentences into parts and to make decisions on spelling in cases where spelling depends on syntax. Both groups improved to a certain degree, but no differences were found between them. Notably, both groups showed marked (and, as may be confirmed by a reanalysis of the data) significant gains on a task which had not been a topic of the instruction, that is, the detection of syntactic ambiguities.

A more recent investigation by Metz (2005) aimed at improving punctuation using an approach whose contents the author describes as "formal-systematic, rule-based grammar instruction" (2005, 56).[1] The instruction featured phases in which students worked in a free, self-determined fashion. Metz notes that a tighter connection between punctuation skills and grammar knowledge seems to have emerged in the experimental classes because the correlation of grammar and punctuation scores increased. She takes this as showing that the experimental students recognized the grammatical base of punctuation which exists in German. However, the experimental students' superiority in the posttests confined itself to grammar knowledge and did not extend to punctuation. Students from schools with high academic aspiration (*Gymnasien*) made greater gains than students from schools with low academic aspiration (*Realschulen*, *Hauptschulen*). In some final interviews that Metz conducted, the latter reported that they lacked pressure to perform better in the experimental instruction.

Summary and discussion. Table 1 offers an overview of the intervention studies reviewed. Effect sizes have been calculated insofar as possible. The effects pertain to the use of grammatical knowledge in metasyntactic tasks (e.g., classification or identification tasks). Effects which pertain only to memorizing declarative knowledge have not been included. If means or standard deviations are not given in the original publication, effect sizes were calculated using the probit formula (see Sánchez-Meca, Marín-Martínez, & Chacón-Moscoso, 2003). If a study reports effects for several achievements, the effect sizes have been aggregated; in such cases, the correlation between the achievements was assumed to be .70. Effect sizes (*d*) and their standard errors (se_d) have been adapted using an approach which takes into account the clustering of the data (Hedges, 2009). This means that the loss of information which results from the assignment of whole classes instead of individuals to the experimental or the control condition is factored in by inflating the standard errors of the effect sizes. The intraclass correlation which is needed to do the calculation has uniformly been set to .20. This conforms to the median of the intraclass correlations in grammar tests which are reported in relevant publications (Helmke & Hosenfeld, 2007; Keller & Moser, 2012; Lehmann & Peek, 1997; Studer & Berger, 2013) or which may be estimated from their data (Claus-Schulze, 1966).

[1] *The original German-language citation has been translated into English by the author of the present text. The same pertains to all other citations of German texts hereafter.*

Table 1. Intervention studies aimed at promoting students' grammatical knowledge

Sample	Grade	Topic of instruction	Methods compared	Number of students (classes)		Pretest	Effect size (standard error)	
				exp.	control		immediate	delayed
Altendorf, 1969, exp. I	5	causal and conditional adverbials	contrasting vs. isolated	109 (4)	109 (4)		.45 (.40)	.04 (.40)
Altendorf, 1969, exp. II	5	infinitival and participial groups	contrasting vs. isolated	63 (2)	65 (2)		.40 (.41)	.38 (.41)
Döhmann, 1977	adults	lexical categories, sentence structures	generative vs. traditional	20 (1)	20 (1)		.38 (.49)	
Friedrich, 1970, exp. I	5	causal and conditional adverbials	contrasting vs. isolated	41 (2)	37 (2)			
Friedrich, 1970, exp. II	5	causal and conditional adverbials	contrasting vs. isolated	37 (2)	45 (2)			
Friedrich, 1970, exp. III	5	infinitival and participial groups	contrasting vs. isolated	93 (5)	55 (2)		.15 (.41)	.41 (.41)
Herrmann, 1974	1-2	sentence types, word classes, sentence parts	concept formation vs. traditional	62 (2)	63 (2)	no	1.25 (.49)	.85 (.48)
Jantos, 1971	2	inflection morphology, word classes, sentence parts	concept formation vs. traditional	281 (10)	351 (13)	inconclusive	.76 (.20)	.53 (.18)
Kramarczyk & Wazel, 1968	4-6	inflection morphology, word classes, sentence parts	programmed learning vs. traditional	60 (2)	552 (20)	no	.34 (.49)	
Metz, 2005	8	finite verbs, subordinate clauses, punctuation	formal grammar and autonomous learning vs. traditional	137 (7)	105 (5)		.61 (.29)	.89 (.29)
Meyer, 1969	5	sentence parts	operational vs. traditional	25 (1)	25 (1)		.41 (.68)	
Ploog, 1981	5	verbs	operational vs. traditional	69 (3)	40 (2)	inconclusive	1.52 (.47)	
Schößler, 1981	4	sentence parts	concept formation vs. traditional	25 (1)	100 (4)		10.15 (.97)	
Schumer, 1988, exp. I	2	adjectives	motivation focus vs. no motivation focus	66 (2)	68 (2)	no	.76 (.48)	
Schumer, 1988, exp. II	3	word formation and comparatives in adjectives	motivation focus vs. no motivation focus	61 (2)	56 (2)	no	1.10 (.49)	
Siegert, 1977, exp. I	5	past participles	operational vs. traditional	113 (4)	53 (2)	inconclusive	1.05 (.37)	
Siegert, 1977, exp. II	5	infinitival groups	operational vs. traditional	96 (4)	49 (2)	inconclusive		

Note. Effects are considered long term if they were measured at least four weeks after the intervention. In the case of Schößler 1981, the number of subjects has been extrapolated from the number of classes. In some cases, the number of subjects may vary between test times.

It should be noted that the effect sizes displayed in Table 1 are estimates. In some cases they rely on data which had to be read off from graphic displays (Döhmann, 1977; Jantos, 1971; Metz, 2005; Ploog, 1981; Schößler, 1981). In two studies (Friedrich, 1970; Siegert, 1977), the publications contain only global distributional parameters from which the original data distribution was partly recovered. This was possible in some but not all cases, and it involves uncertainties to a varying degree. Also, one should bear in mind that the statistical power of a meta-analysis based on the given studies is limited, especially the power to detect heterogeneity. Assuming a significance level of α = .05, the power to detect even strong heterogeneity ($\tau^2 = V_d$) in the data is 0.60; this value drops even more in subsequent cases where fewer studies are included. The power to detect medium treatment effects (d = .30) is higher than that to detect heterogeneity; it ranges between .98 and .52 depending on the assumed heterogeneity in the data.

At first sight, Table 1 seems to reveal marked progress in the experimental classes which experienced intense and mainly formal grammar instruction (mean effect d = 1.30, se_d = .61, p = .03 in the immediate posttests according to a random effects model; see Viechtbauer, 2010). However, as the homogeneity statistic shows, effect sizes are not homogeneous ($Q(13)$ = 103.03, p < .01). The inhomogeneity is caused by the inclusion of Schößler's (1981) experimental group which is reported to have made gains of more than 10 standard deviations as compared to its control group in a 14-week instructional period. But even when one excludes the Schößler sample it would be questionable to calculate a mean effect with the purpose of testing hypotheses about the impact of grammar instruction because the statistical power of the homogeneity test is limited. Inhomogeneity seems to persist because the effects were apparently influenced as much by the quality of the pretests as by the quality of the posttests. A significant difference emerges (z = -2.15, p = .03) when one uses metaregression to compare the immediate effects in the samples where the pretest unequivocally attested to the equivalence of starting conditions in the experimental and the control groups with samples where no pretest had been administered, or where the pretest led to an ambiguous result. The mean immediate effect is d = .44, se_d = .16 in samples with unequivocal pretests but d = .90, se_d = .14 in samples with equivocal or no pretests. Both values are significant but they diverge considerably. The effect in groups with equivalent starting conditions remains stable when one considers long term outcomes (d = .46, se_d = .14).

On the whole it seems that the intensified grammar instruction delivered to the experimental classes had some impact with respect to its proximate target, i.e. explicit knowledge of grammatical concepts. Importantly, this does not imply that students' linguistic skills were affected. International research has repeatedly observed an advantage for experimental students over controls in the knowledge of grammatical concepts as well, however this was not accompanied by progress in linguistic skills (see the classic studies by Harris, 1962, and Elley, Barham, Lamb & Willie, 1979). Though the intervention studies considered thus far put emphasis on the promotion of 'knowledge ready for use' (*anwendungsbereites Wissen*), they do

not, after all, resolve the question of whether this type of knowledge will emerge as a result of formal grammar instruction.

Two intervention studies (Jantos, 1971; Herrmann, 1974) examined whether it is possible to instruct children on formal grammatical features as early as in grades 1 and 2. This deliberate precipitation of grammar instruction may be justified if it pays off in later years of schooling. With respect to the long-term effects of such instruction, however, the data are inconclusive. Jantos's experimental students were superior to the control students in their recognition of verbs from the beginning (d = .44). Immediately after the end of the intervention in year 2, they had increased their lead (d = 1.06). By the middle of year 3, however, the difference was the same as it was at the start (d = .37). From a statistic Herrmann displays it appears that at the end of year 1 his experimental students outperformed the control students by far in the recognition of sentence types (d = 1.32). The same was true at the end of year 2 (d = .89). However, by that time the control students had reached precisely the same performance level which the experimental students had reached a year before. Therefore the question remains as to whether the experimental grammar instruction provided the intervention students with an enduring advantage or merely a temporary one.

3.2 Intervention studies aimed at promoting linguistic skills

Overview. Whereas in Anglophone research, the impact of grammar instruction on students' linguistic skills, especially writing, has been a main focus, little has been done by Germanophone researchers to explore this issue empirically. The latter have tended to take for granted that because grammar knowledge underlies skillful writing (Heringer, 2000; Hoffmann, 2000) and appropriate reading comprehension (Glinz, 1993; Heringer, 2001), grammar instruction will automatically foster both. In the present chapter, consideration will be given first to studies on the impact of grammar instruction on writing (Klotz, 1996; Scholz, 1975) and to studies on reading (Blatt, Müller & Voss, 2010; Funke, Wieland, Schönenberg & Melzer, 2013; Hohm, 2007; Szubert, 2015; see also Herrmann, 1974; Döhmann, 1977). Subsequently, some findings about the effects of grammar instruction on spelling and punctuation will be compiled. The section concludes with a discussion of one isolated study on the acquisition of lexical items in grammar lessons (Siebenbrodt, 1966).

Grammar instruction and writing. In two intervention studies done in grades 5 and 7, Scholz (1975) used 'complex' instructional assignments which required students to devise formulations as well as to analyze them. He focused on the elaboration of texts by the use of attributes (i.e., expressions of various types which accompany nouns, modifying them semantically). Pre- and posttests included grammar and writing tasks. In the posttest, the experimental students performed better on both types of tasks than the control students who had been instructed according to the then-current syllabus. Moreover, they used a greater variety of attributes in compositions written after the intervention. Unfortunately, Scholz

does not document his findings clearly. For instance, he does not specify in which period and under which conditions the compositions he evaluated were written.

In a study of grade 6-8 students from high-track schools, Klotz (1996) examined the hypothesis that connecting teaching grammar to teaching writing would be more effective for the development of formulation skills than teaching both in isolation. As a measure of formulation skills he considered the degree to which sentences are 'informationally saturated'. A sentence expresses a thought which, from a semantical point of view, includes components such as several participants, mode and epistemic modality. The more of these components the sentence expresses, the more it is informationally saturated, according to Klotz. The study found that the posttest compositions of the experimental students featured a significantly higher degree of informational saturation than those of the control students.

Grammar instruction and reading. In the instructional experiment in grades 1-2 described above, Herrmann (1974) collected data on reading skills in some of his subjects at the time of the posttest. Students from the experimental classes outperformed by far students from the control classes on reading prosody and on answering comprehension questions. However, Herrmann also states that experimental students were vastly superior in decoding. Thus one must either assume that decoding was fostered by grammar instruction too—which is rather implausible, especially in grades 1-2—or that the reading abilities of the experimental students exceeded those of the control students independently of the instruction.

Hohm (2007) hypothesized that German studies instruction will foster reading comprehension if it contributes to developing language awareness. He conducted an intervention study in grade 7 in which the experimental classes received instruction with a language focus for several months. The instruction covered highly diverse topics such as word classes and language use in social situations. It also included language practice, such as sentence combining and cloze exercises. The pre- and posttests were comprised of reading comprehension tasks and tasks requiring students to analyze language phenomena occurring in text. Hohm summarizes his findings as showing that the experimental students made significantly greater gains than the control students on the language tasks but not on the reading comprehension tasks. A reanalysis shows that the effects were exclusively due to gains made by L2 learners.

Blatt et al. (2010) in grade 5 supplemented an earlier experimental intervention on reading fluency by means of a follow-up program with a focus on grammatical, orthographic and textual analysis. Children who were non-responders to the reading fluency intervention were included in the experimental group of the follow-up study. They received intense individual tutoring by preservice teachers for about six months. The control group consisted of participants in the fluency study who were parallelized according to reading comprehension. In the pre- and posttests, a test of negative cloze and a reading comprehension test were used, both of which were standardized. Blatt et al. found that, compared to the control students, the experimental students caught up in negative cloze but fell off in reading comprehension during the intervention period. In a comment, the authors point out

that the quality of the tutoring delivered by the preservice teachers seems to have varied to a high degree.

Funke et al. (2013; cf. Melzer, 2011) implemented an instructional program including four lessons in grade 6 classes. In the instruction, they used 'direct access tasks' of the type described below (see 3.3.) in which nouns and finite verbs were contrasted, as well as direct access tasks which tapped the contrast of nominal phrases with nominative case and nominal phrases with accusative case. In the pre- and posttests, the study employed freely compiled tasks on reading (reading comprehension; evaluating capitalization in reading) and on spelling (word level spelling; capitalization). The control classes were instructed according to teachers' plans. Funke et al. found no difference between experimental and control students with respect to the development of reading comprehension. With respect to the tasks which examined how capitalization was evaluated in reading, those experimental students who ranked either high or low in the pretest improved compared to the controls; however, students who ranked in the medial group in the pretest fell off.

Szubert (2015), in a small study in grade 3, conducted an experiment in which the children in the experimental group practiced combining word cards into sentences during five lessons of individual training ('anagram task' according to Weaver, 1979). In ensuing dialogues, they were encouraged to try various ways of arranging the words and to check what 'sounds best'. The control children continued to participate in whole-class instruction. In his pre- and posttests, Szubert focused on the grasp of information structure (finding out which version of a sentence best profiles what information is new and what information is old in a given context; see, e.g., Lambrecht, 1994) by using self-conceived tasks. Reading prosody was also tested. The grasp of information structure improved significantly more in experimental children than in control children.

Grammar instruction and spelling. In German, decisions on how to spell may make it necessary to take into account syntactic features. This pertains to the capitalization of nouns and, in some cases, to grammatical morphemes. Moreover, punctuation depends on sentential structure to a high degree. Several of the studies reviewed so far deal with syntactic spelling. Döhmann (1977) found that the adult students in her intervention groups improved in punctuation as well as in orthographically marking a grammatical morpheme (conjunction *dass* 'that' versus relative pronoun *das* 'that'). In the Funke et al. (2013) study, grade 6 experimental students made significantly greater gains than the control students in capitalization, with experimental students improving evenly regardless of their starting conditions. No significant difference was found in general spelling. Jantos (1971) notes in passing that the effects of his grammar instruction program in grades 2-3 on students' capitalization skills seem to have been modest. Unfortunately, the figure in which he displays his results pertaining to capitalization is apparently miscalibrated. No specific effects of grammar instruction on punctuation were found by Metz (2005) in grade 8, though the experimental instruction was tailored to foster punctuation.

Grammar instruction and vocabulary learning. One of the sources of vocabulary learning in school-aged children is independent reading in which word meanings have to be derived from the context. Nagy (2007) hypothesizes that deriving word meanings from context requires what he calls 'syntactic awareness'. This might mean that vocabulary learning is a case for grammar teaching. Whether grammar instruction indeed fosters word learning does not seem to have been explored in research from German-speaking countries. However, the acquisition of vocabulary was one focus of a field study on grammar instruction in grades 6 and 8 by Siebenbrodt (1966). Siebenbrodt was concerned with whether the linguistic forms which teachers deliberately introduce in grammar lessons are adopted by students in the language they actually use. She considered grammatical forms such as verbs in conjunctive mood, but also lexical forms such as uncommon words and complex expressions, and tried to determine whether students took up the forms by evaluating diverse texts written by them in the course of their schoolwork, such as homework, school exercises, and compositions. Siebenbrodt found that students rarely adopted new forms that teachers had brought to their attention in grammar lessons. If forced to do so by an assignment, they often seemed to struggle with them. In an instructive case reported by Siebenbrodt, the phrase at stake was *sich einer Sache bewusst sein,* 'to be aware of something'. It is a verbal expression which takes two arguments. The first serves as a Theme (the noun phrase which says what the awareness is about; it is represented by 'of something' in English and by *einer Sache* in German). The second argument serves as an Experiencer (the noun phrase which says who experiences the process of becoming aware; it is represented by the subject and, in German, resumed by the reflexive pronoun *sich*). The Theme argument is in the genitive case, which is exceptional in modern German. Siebenbrodt found that in a grade 6 class, 77% of the students seemed not to be able to use the expression with its complete verbal argument frame after it had been the topic of a grammar lesson. In most cases, they failed to include the reflexive pronoun which resumes the Experiencer argument. Even worse, some of them replaced the verbal expression with a noun (*Bewusstsein*) or with an adjective having an argument-absorbing suffix (*bewusstlos*). It seems that their problem was assigning an argument structure to the expression. Siebenbrodt found some further cases in which students' reluctance to adopt new forms seemed to arise from difficulty understanding their grammatical features. In most of these cases, argument structure was involved.

Summary and discussion. Table 2 summarizes the results of studies on the impact of grammar instruction on linguistic skills based on the same methodology as described in 3.1. At variance with the general procedure (see comment on Table 1), the effect size for the Funke et al. 2013 study was calculated using the odds ratio formula because scores were obviously not normally distributed in this study. Recall that the statistical power to detect heterogeneity is very limited in these cases.

Table 2. Intervention studies aimed at fostering students' linguistic skills

Sample	Grade	Topic of instruction	Methods compared	Number of students (classes)		Pretest	Effect size (standard error)	
				exp.	control		specific	unspecific
Writing								
Klotz, 1996	6-8	adverbials and their functions in texts	integrated vs. isolated	30 (3)	30 (3)	not used	.56 (.44)	
Scholz, 1975, grade 5	5	attributes and their functions in texts	communicative vs. traditional	30 (1)	30 (1)		.26 (.68)	
Scholz, 1975, grade 7	7	attributes and their functions in texts	communicative vs. traditional	29 (1)	29 (1)		.82 (.68)	
Reading								
Blatt et al., 2010	5	grammatic, orthographic, textual analysis	individual tutoring vs. no training	32 (16)	39 (16)			-.32 (.23)
Döhmann, 1977	adults	lexical categories, sentence structures	generative vs. traditional	20 (1)	20 (1)		.23 (.69)	
Funke et al., 2013	6	syntactic categories, case	direct access vs. school-type instruction	119 (6)	137 (6)		-.02 (.28)	-.10 (.28)
Herrmann, 1974	1-2	sentence types, word classes, sentence parts	concept formation vs. traditional	62 (2)	63 (2)	no	1.98 (1.20)	1.05 (1.30)
Hohm, 2007	7	word classes, language use	language awareness vs. school-type instruction	40 (2)	44 (2)		.04 (.49)	-.27 (.49)
Szubert, 2015	3	building sentences	anagram training vs. school-type instruction	5 (1)	7 (1)		.87 (.56)	
Spelling and punctuation								
Döhmann, 1977	adults	lexical categories, sentence structures	generative vs. traditional	20 (1)	20 (1)		.00 (.69)	
Funke et al., 2013	6	syntactic categories, case	direct access vs. school-type instruction	119 (6)	137 (6)		.32 (.28)	.25 (.28)
Jantos, 1971	2	inflection morphology, word classes, sentence parts	concept formation vs. traditional instruction	233 (10)	373 (13)	inconclusive		
Metz, 2005	8	finite verbs, subordinate clauses, punctuation	formal grammar and autonomous learning vs. traditional instruction	137 (7)	105 (5)		.11 (.28)	

Note. Effects are considered to be specific if those linguistic achievements are affected which draw directly on the grammar knowledge taught. Unspecific effects pertain to overall writing quality, reading comprehension or general spelling. In some cases, the number of subjects may vary between test times.

In the last two columns of Table 2, a distinction is made between specific and unspecific effects. An unspecific effect is taken to be present if a linguistic achievement has been globally affected (e.g., overall quality of writing, global text comprehension in reading, or word spelling in general). A specific effect is supposed to exist if a component of a linguistic achievement which draws on grammar knowledge was affected by the instruction (e.g., use of grammatical forms in writing, interpreting grammatical morphemes in reading, or syntactic spelling). As can be seen from Table 2, no unspecific effects of grammar instruction on linguistic skills have been observed (for reading comprehension, the mean effect is d = -.22, se_d = .17, p = .19, homogeneity statistics $Q(3)$ = 1.33, p = .72; for writing and spelling, no mean effect can be computed). With respect to writing quality, the international literature generally concurs that no specific effects of grammar instruction exist (Graham & Perin, 2007) though recently evidence has been proffered that grammar instruction focusing systematically on writing problems will yield better results (Myhill, Jones, Limes & Watson, 2012). With respect to reading comprehension, the lack of unspecific effects does not come as a surprise given that what reading comprehension tests measure is to a large extent determined by factors not related to grammar, such as domain-specific knowledge, use of reading strategies, and decoding skills (Perfetti, Landi & Oakhill, 2005). Under this condition it is ex ante unlikely that reading comprehension can be developed by grammar instruction. Ultimately, there is no compelling reason to expect that grammar instruction would promote the spelling of words in cases where orthography does not depend on sentential context.

Though the values for specific effects listed in Table 2 never aggregate to significant effects, it might be worthwhile to further explore specific effects of grammar instruction on linguistic achievements. As for writing, significance with regard to linguistic achievements is missed only barely (d = .55, se_d = .32, p = .09; homogeneity statistics $Q(2)$ = .34, p = .84). International research found that is possible to influence the linguistic features of texts written by students by means of grammar instruction which is delivered in tight connection with formulation problems (see Rogers & Graham, 2008). As for reading, there is no significant specific effect (d = .20, se_d = .21, p = .34; homogeneity statistics $Q(4)$ = 4.34, p = .36). However, in some studies, achievements which may be supposed to be components of reading comprehension were shown to be affected by instruction with a grammar focus (detecting syntactic ambiguity, grasping information structure). The list of the components considered specific is highly diverse and was compiled ad hoc; beyond that, the effects reported seem to be more heterogeneous than the homogeneity statistic suggests. Research might make progress if it explores in a principled, theory-based fashion any cases in which grammar learning might affect reading comprehension. With respect to spelling and punctuation, the aggregated effect is nonsignificant (d = .20, se_d = .19, p = .29; homogeneity statistics $Q(2)$ = .37, p = .83). Nevertheless, it is conceivable that grammar learning has effects on the orthographic marking of syntactic structure. Note that the zero effect reported for the Döhmann (1977) study shows that the 'traditional' group improved to the same

extent as the 'generative' group. Taken together, however, both groups made significant gains which manifest in a combined pretest-posttest effect size of d = 1.15, se_d = .35. Thus, the combined effect size computed for all studies in Table 2 underestimates the effect of grammar instruction on syntactic spelling. The picture is less encouraging with respect to punctuation. Instructional approaches to punctuation with a focus on grammar have not been proved to be very effective so far (Erb, 1982; Sappok, 2011). The arrival of the study with adults (Döhmann, 1977) at another result may be because the subjects in this study deliberately set out to improve their spelling in the course, as evidenced by a final interview. One might speculate whether, in order to successfully punctuate based on grammatical knowledge, students must draw on a global process of syntactic analysis which is different from the local grasp of syntactic features related to capitalization and the spelling of grammatical morphemes.

Strangely enough, virtually no research has been conducted whether or how grammar learning at school affects learning how to derive word meanings from context. This seems to apply to the international literature as well as to Germanophone research.

3.3 Observational studies

Overview. Studies are labeled 'observational' if they restrict themselves to exploring the scope and quality of grammar knowledge without intervening in instruction. Their main objective has been to determine the extent to which students grasp syntactic features. Some of them have focused on students' grammar concepts, drawing on verbal data such as questionnaires, interviews, and classroom logs (König, 1992; Mesch & Dammert, 2015; Müller & Tophinke, 2015; Peyer, 1997; Riehme, 1972; Rose, 1997; Schübel, 2000; Schuttkowski, Rothstein, Schmitz & Gräsel, 2015; Spies, 1989). Other studies drew on test data to investigate students' grammar knowledge (Claus-Schulze, 1966; Funke, 2005; Habermann, 2013).

Studies using verbal data. In a small but influential inquiry, Spies (1989) concerned herself with the ideas the students of a third grade class voiced about what nouns, verbs and adjectives are. She logged the reasons children gave for why they considered a word to belong to one of these word classes. According to her, children assigned words to word classes predominantly based on semantic criteria, and, when doing so, relied on a pictorial rather than a formal way of thinking. This assumption seems at first sight to be supported by numerous casual observations described in the literature (e.g., Andresen, 1996; Jantos, 1971; Röber-Siekmeyer, 1999; Valtin, 1988; Weingarten, 2001).

However, Mesch & Dammert (2015), in a study of the concept of verb in grade 4 children, found no correlation between children's ability to detect semantically peripheral (non-agentive) verbs and their propensity to argue upon questioning that verbs may be discerned by their reference to actions. So, how children reason about syntactic categories does not necessarily reveal how they detect them when words occur in an utterance. Moreover, according to casual observations described

in the literature, children assigning words to word classes seem to be misled by purely formal considerations as much as by semantic considerations (Eichler, 1992; Funke, 1995a; Röber-Siekmeyer, 1999).

The knowledge of more advanced students regarding word classes was investigated by Riehme (1972) in grades 5-9, König (1992) in grade 5, Rose (1997) in grades 5-7, Schübel (2000) in grade 6 and Müller & Tophinke (2015) in grades 5-9. In a study on capitalization skill, Riehme had students explain how to recognize nouns in texts. Students from grades 5-6, where formal grammar is a more prominent topic of instruction than in later years, frequently referred to grammatical criteria. Older students responded in a more implicit way, often referring to semantic criteria. In König's and Rose's samples, students more easily identified nouns than verbs and adjectives. It might be added that Jantos (1971) observed the same thing in his sample. However, Muhr (1995) and Funke (1995b), considering a less explicit task, observed that verbs were rarely miscapitalized in students' writing when they appeared in finite form. This suggests that the difficulty in identifying verbs manifests only when they have to be classified apart from their context. Müller & Tophinke ascertained that students were hardly able to explicate their knowledge about word classes in a conceptually satisfying fashion. This result echoes the findings of other studies on grammar knowledge (Peyer, 1997 with respect to knowledge about sentences in grade 9; Schuttkowski et al., 2015 with respect to knowledge about syntactic categories and tense in grades 5-9).

Studies using test data. The most comprehensive descriptive study on students' grammatical knowledge was conducted by Claus-Schulze in the former German Democratic Republic (1966). She assigned grade 5-9 students tasks in which they had to either classify linguistic units according to syntactic features ('analytical tasks') or sketch formulations according to preset specifications ('synthetical tasks'). Moreover, she (hand-)recorded grammar lessons in a multitude of classes, most of which completed her tests too. Claus-Schulze found marked differences in performance between the classes. Based on her classroom observations, she explained these differences by distinctions in the consistency teachers showed in challenging their students. In particular, she addressed the extent to which the solutions students proposed in the tests attest to reliable and substantially founded grammatical knowledge. It becomes evident that she is skeptical about that given her characterization of students' grammar knowledge as "insufficient" (1966, 78).

The most instructive component of Claus-Schulze's study is the analysis she performed on diverse types of test tasks. In tasks which required identifying the subjects and predicates of sentences, Claus-Schulze arrived at a conclusion analogous to that of Jantos (1971): Students frequently marked subjects without marking the related predicate and vice versa. Moreover, they classified main verbs as predicate parts more often than auxiliary verbs. In this case, it appears that they focused on the features of isolated words rather than on their role in the given sentence. This suggests a "word-by-word procedure" (Claus-Schulze, 1966, 150) which may lead to correct solutions although it is left open whether students possess genuine syntactic knowledge. Claus-Schulze offers an impressive example

from the test materials she used in grade 5. Students were required to mark the subjects and predicates of sentences by highlighting them. One of the sentences reports on the vacation spot where some children spent their summer holidays. Claus-Schulze's analysis focuses on the predicate. She expected students to mark the words which are underlined in the following sentence.

> *Dort werden sie mit Vergnügen wandern und baden und erholt zurückkommen*
> (There they will hike and swim with pleasure and come back home refreshed)

The verbal chain is headed by the auxiliary *werden* 'will'. It is supplemented by three infinite verb forms, *wandern* 'hike', *baden* 'swim' and *zurückkommen* 'come back home' which are all related to the auxiliary in the same way. From the data Claus-Schulze includes, one can calculate the counts displayed in Table 3.

Table 3. Marking of parts of the verbal chain by grade 5 students in a sentence stemming from the test materials used by Claus-Schulze (1966)

cases where *zurückkommen* has been marked				cases where *zurückkommen* has not been marked			
		baden has been marked				*baden* has been marked	
		yes	no			yes	no
wandern has been marked	yes	109	6	*wandern* has been marked	yes	106	9
	no	7	33		no	4	31

When one compares the left- and the right-hand panels of Table 3, it becomes evident that about 75% of the students correctly marked *wandern* and *baden* as parts of the predicate. Both cases are tightly connected: In most cases where students mark *wandern*, they mark *baden* as well, and vice versa. This leads to the conclusion that students who marked these words grasped the fact that both perform the same role in the sentence. Precisely the same role is performed by *zurückkommen* too. However, students seem to treat *zurückkommen* as a completely different case. Only 50% of them mark it, and whether they mark it or not is independent of whether they mark the other verbs. This leads to a diametrically opposed conclusion, namely, that students disregarded the word's role in the sentence and just took it in isolation. Claus-Schulze infers that when working on metalinguistic assignments, students seem to grasp syntactic features in a way which is "neither essentially reliable nor essentially unreliable" (1966, 296).

In a study in grades 5-7 on the understanding of syntactic categories, Funke (2005) designed tasks to explore the extent to which students grasp the contrast of nouns and finite verbs independently of applying grammatical operations and using grammatical terminology. Students were given four sentences and asked to tick the sentence which is 'different from the other ones', as illustrated below.

In that little Celtic village there live people who oppose the rule of Caesar.

- *One of these rebels fights the Roman officers.*
- *One of these rebels constantly resists the Romans.*
- *One of these rebels vividly combats the Romans.*
- *One of these rebels against the Roman officers.*

In the case given, the fourth sentence must be picked out because the unit *rebels* has to be taken as a finite verb in this case, whereas in the other cases it functions as a noun. Tasks of this type in which a critical unit appears several times but differs in its syntactic role in one case compared with the other cases may be labeled 'direct access tasks'. The students were given 20 such tasks. After they worked through them, they were presented with 20 tasks in which they had to decide whether words which appeared in a text were verbs or not. These tasks, which used grammatical terminology to make the assignment clear, may be labeled 'school-type tasks'. Only those students whose score in the direct access tasks ranged above chance level were able to reliably solve the school-type tasks. The reverse was not true. The distribution of hits in the direct access tasks was bimodal, featuring its maximum at the highest possible score of 20 and another local maximum at a score of 11. What is remarkable about the lower maximum is that it exceeds the value one should expect if the students in the lower tail of the distribution responded by chance (which amounts to 5). That is, even these students in their majority appear to have had access to syntactic information relevant to the contrast at stake. However they seem not to have been able to reliably access this information when working on the task.

In the direct access tasks, the difficulty of items did not seem to depend on semantic features of the critical unit. It was .71 when the critical unit, in its noun reading, was a concrete noun, and .72 when it was an abstract noun. Likewise, it was .72 when the critical unit, in its verb reading, featured an agentive meaning, and .71 when it featured a non-agentive meaning. By contrast, in the school-type tasks the common finding was corroborated that agentive verbs are easier for students to identify than non-agentive verbs (difficulty .91 versus .47). This means that when one looks at the school-type tasks one may be fortified in the widely held belief that students' attention is attracted by semantic features at the cost of syntactic features. In the direct access tasks another picture emerges. As the author proposes, the bimodal score distribution in the direct access tasks finds its explanation if one assumes that when students work on metasyntactic tasks, syntactic features compete not with semantic features but rather with other syntactic features. A maximum in the range of 11 (more precisely, 10) points would be expected if students in the lower tail of the distribution understand that only two readings (a verb and a noun reading) are possible but have difficulty maintaining the appropriate reading because the other one interferes. In this view, two syntactic options compete in students' minds, one in which the critical unit features as a noun and one in which it features as a finite verb.

Supplementing observational studies done in schools, a much-cited study by Habermann (2013) looked at the grammar knowledge of university students who

had just left school. 357 undergraduates worked on a test which had been devised by the Bavarian authorities for grade 8 students. It examined grammar knowledge in addition to reading, writing and spelling abilities. The undergraduates outperformed the grade 8 students in all domains of the test with the exception of its grammar part, in which the grade 8 students were more successful.

Summary and discussion. The studies which used verbal data to explore students' concepts of grammatical categories only rarely featured methodological rigor. Their main finding has been that students, even secondary students, had great difficulties giving conceptually satisfying definitions of grammatical concepts. An implicit assumption in these studies is that students consequently do not grasp the syntactic features the concepts denote. However, students being unable to define the concept of verb did not preclude their being able to identify instances of verbs in a text (Mesch & Dammert, 2015). In addition, when students are asked to explicitly define grammatical concepts, their definitions seem to become sloppier after grammar instruction has ceased to be prominent in instruction even though there are no indications that the grammar knowledge needed to identify syntactic construction actually decreases (Riehme, 1972).

Claus-Schulze (1966) concluded that the grammar knowledge of grade 5-9 students was, on the whole, 'insufficient'. What struck her most was that students seemed to grasp the syntactic structure of sentences only on a case-by-case basis. As an example, she considered a sentence about the vacation spot of a group of children.

There they will hike and swim with pleasure and come back home refreshed.

Claus-Schulze wondered why students recognized that 'hike' and 'swim' are parts of the predicate but seemed to treat 'come back home' as a completely different case even though it has the same syntactic status. The solutions offered by the students may be understood if one considers that the adverbial 'there' which features in the initial position of the sentence may, for semantical reasons, only be related to 'hike' and 'swim', not to 'come back home'. You can hike at a vacation spot and you can swim there, but you cannot come back home there. If anything, you can come back home *from* there. Being located in initial position, the adverbial has scope over the sentence. From this it follows that, if one uses adverbial scope instead of punctuation to delineate the sentence, 'hike' and 'swim' belong to the same sentence whereas the syntactic relation of 'come back home' to this sentence is left unspecified. The solutions students offered do not necessarily show a lack of sensitivity to syntactic structure because syntactic structure, as accessed by students, must not be equated with sentence structure which is the topic of school grammar.

Funke (2005) found access to the syntactic category contrast of nouns and finite verbs to be widespread in secondary school students when probing for it by means of a direct task. Moreover, this access neither interfered with semantic features of the verbs and nouns used in the task nor did it depend on students' ability to explicitly classify words as nouns or verbs. However, some students were unable to

reliably detect the syntactic contrast. An explanation for this might be found if one assumes that students, when they detect the contrast in a set of sentences, rely on syntactic information which they generate spontaneously when processing the sentences. If so, they will be left in an undetermined state of mind whenever this syntactic information has not been spontaneously generated, and this will lead to errors comparable to those Claus-Schulze observed. This leads to the hypothesis that the syntactic information one generates during linguistic processing is different from the evidence one accumulates by analyzing sentences. The latter may be reliably found in a task-driven fashion if one proceeds systematically, whereas the former may emerge reliably only if it one habitually relies on a data-driven process.

3.4 Large-scale studies

Overview. Large-scale studies are a subtype of observational studies aimed at measuring the distribution of competences in a given population on the basis of a representative sample (or even the whole population). Grammatical knowledge has been included in two large-scale studies restricted to German-speaking regions (LAU study, DESI study), and it has also been a target domain of some educational monitoring surveys (IQB studies, VERA studies, 'Check 5' project, 'Lernstandserhebung Zürich'). In all cases, data were evaluated using a one-dimensional item-response model. In what follows, we will take only a brief look at the large-scale studies by first considering their main findings and then examining the competence models which have been devised based on them.

Main findings. In longitudinal surveys, the LAU study (Lehmann, Peek & Gänsfuß, 1997, 1999; Lehmann, Gänsfuß & Husfeldt n. d.) and the DESI study (Eichler, 2007a) mixed tasks tapping grammatical knowledge with tasks which required assessing linguistic correctness and stylistic adequacy. They arrived at rather diverse results on students' learning progress: whereas LAU found only modest gains from grade 7 to grade 9 (pre-post effect size d = .24 in two years), DESI reports substantial gains in grade 9 alone (d = .35 in one year). Similarly, surveys aimed at educational monitoring do not converge as to the extent to which students acquire the intended competences: whilst Helmke & Hosenfeld (2007) affirm that the set standards are met in the domain of analyzing language, Bremerich-Vos & Böhme (2009) are more skeptical, and KMK (2014) does not make a firm statement on this question.

A commonality of the LAU and DESI studies is that both found the achievement gap between high-track schools (e.g., *Gymnasien*) and low-track schools (e.g., *Haupt- und Realschulen*) to be greater on tasks requiring the analysis of language than on other tasks related to German studies (such as reading comprehension). Data suggest the same effect prevailed in a survey done with Swiss 11th-graders (Keller & Moser, 2012).

Competence models. Based on the data collected in large-scale studies, attempts were made to arrange tasks according to their difficulty. VERA (Isaac, Eichler & Hosenfeld, 2008) and IQB (Bremerich-Vos & Böhme, 2009; KMK, 2014)

present models which, though differing in details, converge in assuming that tasks on analyzing language are all the more difficult as they require students to attend to whole sentences instead of single words. This fits with the casual description of what made tasks difficult offered in the LAU study (Lehmann et al., 1999). A different view, however, may be abstracted from a study intended to pilot the harmonization of educational standards in Swiss cantons (Konsortium HarmoS Schulsprache, 2010). In that study, it is implicitly assumed that an item covering each type of task may be found at any competence level. This also seems to be the case in a study based on data from the 'Lernstandserhebung Zürich' project (Moser, Buff, Angelone & Hollenweger, 2011), which traced the development of Swiss primary school children from grades 1 through 6.

Establishing competence models in the sense of arranging tasks post hoc according to their difficulty is a purely descriptive endeavor. Competence models which have a more theoretical foundation have been presented by researchers involved in DESI (Eichler, 2007b) and the related VERA project (Isaac et al., 2008; see also Isaac, Hochweber & Eichler, 2010; Isaac & Hochweber, 2011). DESI considers metasyntactic achievements to be a manifestation of language awareness, which is taken to be the same thing across diverse linguistic domains (phonology, morphology, lexicon, syntax, pragmatics). Note that this is not self-evident because metasyntactic achievements might require specifically syntactic learning processes instead of being the result of becoming aware about language in general. Eichler (2007b) presents a regression model which strives to predict task difficulties in the domain of language awareness. In order to explain the data, domain-specific parameters had to be included above general difficulty parameters in the model. Isaac et al. (2008) try to predict task difficulties assuming ten difficulty parameters. Most task difficulties could be predicted using a regression model but significant deviations were observed in two cases: First, the model overestimated the difficulty of tasks which require students to decide on the sociopragmatic appropriateness of formulations, and second, the model underestimated the difficulty of tasks requiring students to correct syntactic errors.

Summary and discussion. The inclusion of grammar knowledge in large-scale surveys is specific to the European tradition in general (e.g., Sijtstra, van der Schoot & Hemker, 2002). One-dimensional item-response theory (IRT) modeling figured prominently in this type of research. It is important to understand that mixing grammar tasks with other language awareness tasks in a one-dimensional metric carries with it an implicit assumption of homogeneity. That is, it implies that solving grammar tasks draws on the same cognitive processes as solving language awareness tasks in general. Eichler (2007b) and Isaac et al. (2008) argue for this hypothesis by pointing to regression models which predicted task difficulties based on a unitary set of predictors. These models, however, did not yield correct predictions for all tasks. Moreover, the large-scale studies led neither to a consistent estimation of the learning progresses made by students nor to a shared picture of what makes a grammar task difficult. Both findings might result from a lack of homogeneity in the tasks used though there may be other explanations as well.

A reliable finding may be identified in diverse large-scale studies' concurrence that the achievement gap between high-track and low-track schools' students is greater in tasks pertaining to metalinguistic knowledge and metalinguistic skills than in tasks pertaining to other realms of German studies. It would be interesting to determine whether this results from differences between the students enrolled in the various school types or rather from differences in school curricula, which in turn might be traced back to differences in instructional cultures.

3.5 Classroom discourse and grammar instruction

Overview. Traditional teaching practice in classroom discourse in general has been the subject of much pedagogical research, so the first of three steps when reviewing this topic will be to consider how discourse in grammar lessons has been investigated along these lines (Herrlitz, 1990; Roeder & Schumer, 1976; Wolt, 1974). Wolt (1974) holds that communicative restriction in grammar lessons results not only from teachers' adherence to traditional teaching practices but also from the fact that it is difficult for children to "consider language as an object" (126). According to this assumption, communication problems in grammar lessons are specifically tied to their topic. How this might come about surfaces in microanalytic studies following an interpretive procedure which will be presented in a second step, organized thematically (Boettcher, 1994, 1999; Bremerich-Vos, 1995, 1996; Brünner, 1982; Kleinbub, 2012, 2014; Simmel, 2007). The third step reviews studies which set out to assess the cognitive quality of discourse by using coding schemes (Claus-Schulze, 1978, 1982; Stahns, 2013, 2014).

Traditional teaching practices in grammar lessons. Older educational research characterized communication in school lessons as rigid and restricting students' opportunities for action. This manifests, among other things, in the predominance of an IRE (Initiation-Response-Evaluation) pattern (Mehan, 1994). In some studies, classroom discourse in grammar lessons was compared to that in other school subjects. Wolt (1974), in five audiotaped grammar lessons in grade 5 classes, found no single student contribution comprising more than two sentences. This contrasts with the majority of the other German studies lessons she studied (18 lessons, of which 11 featured student contributions of more than two sentences). Wolt also reports that the proportion of 'bound' student utterances (i.e., utterances which were restricted to conform to a given linguistic pattern) was higher in the grammar lessons. Herrlitz (1990) compared transcribed grammar and literature lessons from five European countries, one of which was Germany. He found that the IRE pattern "seems to be the dominating turn taking structure in … grammar instruction" (1990, 13). Roeder & Schümer (1976) studied instructional communication in an intervention study in grade 3 in which twenty lessons were audiotaped, five of them dealing with grammar. They did not observe a marked communicative restriction in the grammar lessons as opposed to the other lessons; however, communication in the grammar lessons appeared to be erratic and lacked thematic conti-

nuity. One might add that communicative restriction and rigidity may be less prominent in grammar instruction nowadays than they were in the past.

Microanalytic studies. As noted, it is often reported that children tend to argue on a semantic basis when reasoning about word classes. Bremerich-Vos (1996) documents the transcript of a grammar lesson in grade 5 in which the teacher strives to guide children from a semantic to a grammatical concept of verb. At two points, this lesson comes to a complete standstill with the children and the teacher seemingly confused. In one case, this results from students' apparently innocuous attempt to establish that the words *laufen* 'to run' and *fühlen* 'to feel' are not nouns but verbs. To achieve this, the students apply a criterion frequently taught to primary students, namely, that a noun refers to an object one can see or touch. Bremerich-Vos cites the following exchange (1996, 216; S is 'student' and T is 'teacher').

26 S: What you do you can't touch.
27 T: Can you explain?
28 S: If I run, somebody from the class cannot touch me.
29 T: Yes, that's true. Doesn't help much yet. Pierre?
30 S: To run is no thing.
31 T: Are we concerned with things?
32 S: No.
33 T: (kindly): Don't let yourself be misguided. Thomas?
34 S: first part incomprehensible, seems to refer to the verb *to touch*): Well because eh if someone feels something then the other one cannot see it. (incomeprehensible)
35 T: Now, what is it the other person cannot see?
36 S: That he feels something.
37 T: Yes, so if I am just feeling something then as a rule someone else cannot see it. (Pause) Where does it operate, the feeling? Where does it happen?

Bremerich-Vos notes that what is happening here seems to be the doing of philosophy with children rather than grammar instruction.

Grammar lessons which deal with sentence constituents and their functions have more frequently been analyzed than lessons on word classes. Generally, children are taught that sentence constituents are made up of words which stay together when they are shifted from one position in the sentence to another. Subsequently they are instructed to determine the functions of the constituents by asking sentence constituent questions (*Satzgliedfragen*). The most common of these questions are *Wer oder was?* 'Who or what' for the subject and *Wen oder was?* 'Whom or what' for the direct object. It has often been reported that children find it hard to understand this procedure. Boettcher (1994) renders a scene in a grade 5 class in which students make proposals as to which sentence constituent question might be appropriate for a given direct object expression. In seven successive attempts, they try various questions before one of them arrives at a solution which satisfies the teacher (*Wen?*). This occurred although the very first proposal offered by a student was, in substance, correct (*Was?*). The teacher, however, wanted the *Was?* question to be accomplished by *Wen?* The reason is that this makes the accusative case of the direct object become apparent. The teacher's

expectation led to a student behavior which Granzow-Emden describes as “wildly quibbling” (1999, 173). It seems that students did not understand why the teacher continued to solicit proposals even though a question which obviously fits well had already been formulated.

A different type of problem with the asking of sentence constituent questions becomes apparent in a scene drawn from a lesson in grade 4 which Kleinbub (2012, 2014) analyzes. In it, the class works with the sentence ‘This is a book’ (in its German version), which relates to a story with which the children are familiar and in which the main character, Lippel, reads a book. The children wonder which question might align with ‘a book’. The answer the teacher expects is ‘Who or what is this?’ indicating that she takes ‘a book’ to be the subject (which is misguided, as Kleinbub notes in passing). The class accumulates five proposals, the most interesting being ‘Whose book reads Lippel?’ and ‘What does Lippel read? The book’. The key feature of both formulations is that the target expression ‘book’ is included in the question.

Precisely the same thing is found in a lesson in grade 3 which Brünner (1982) analyzed. When asked to formulate the constituent question for the subject of the sentence ‘You know aquanauts’, children first proposed ‘Who or what do you know? Aquanauts’, and then ‘You know aquanauts.’ The author notes that in this moment the lesson ended up “in complete confusion for teacher and students” (1982, 141).

Finally, the problem also surfaces in an incident which Bremerich-Vos (1995) documents. A grade 6 student tries to pose the question appropriate to the direct object of the sentence ‘In the afternoon, I attended a circus performance and experienced something very special.’ He proposes ‘Whom or what very special experienced’ and ‘A circus performance. Who or what I have.’ Neither he nor an adult observer who happens to come around are able to arrive at a conclusion regarding the right question. When students include the expression which is the target of the question as part of the question, this suggests that sentence constituent questions challenge students not only because it is hard to understand why one should pose them but also because they put high demands on students’ ability to formulate deliberately.

The work reviewed so far might seem to suggest that microanalytic studies have found only deficits in the grammatical knowledge and understanding of students; this, however, is not the case. Boettcher (1999) discusses a lesson in a grade 5 class dealing with a topic which is, as the author notes, highly subtle: namely, the distinction between adverbials and prepositional objects. However, that lesson included no misperformances of the type illustrated above. This is true even though students had to classify sentence constituents very frequently. It suggests that the wall between understanding and not understanding in grammar lessons is less impervious as it seems as long as one obstinately zeros in on trying to break it.

Evidence for understanding is also reported by Simmel (2007), who focused on how students explain grammatical phenomena to each other. In her study she used the direct access tasks described above (Funke, 2005). Simmel presented the tasks

on a computer screen and had two students at a time from grade 5 or grade 8 classes work collaboratively on them. In the materials she collected, repeatedly a sudden solution was found after a while of grappling with the task. In the typical case, the solution stood in no relation to the foregoing argumentation but rather thwarted it.

Cognitive quality of discourse. Claus-Schulze (1978, 1982) recorded (in writing) German lessons of 45 teachers. She collected compositions from a partial sample of the students; from some of them she also collected test data on grammar knowledge. The lessons recorded dealt not with grammar but with literature; however Claus-Schulze analyzed them with respect to the question of which results of grammatical learning might show up. Teachers' and students' utterances were coded according to their linguistic complexity and their level of cognitive aspiration. The study found that in lessons where teachers' utterances reached a high level of cognitive aspiration, the same was true of students' utterances. Moreover, students' utterances were linguistically more complex in these lessons, comprised more metalinguistic comments on the texts serving as the topic of the lessons, and more frequently indicated that students monitored their linguistic behavior or that of their classmates. Claus-Schulze emphasizes that, surprisingly, the linguistic complexity of students' utterances was not correlated to their knowledge of school grammar as evidenced in the grammar test.

In a recent investigation on the 'opportunity structures' of grammar instruction (i.e., the learning opportunities students are allowed), Stahns (2013, 2014) analyzed seven videotaped grammar lessons from grades 5 and 7. He focused on the question of the extent to which instruction contributed to the cognitive activation of students. Teacher actions which are estimated as cognitively activating in the literature were rarely found in the lessons recorded. This pertains to open questions (fewer than 20% of all teacher questions) and to questions aiming at deep reasoning (fewer than 10% of all teacher questions). Teacher solicitations which required using specifically scientific methods were rare, too. As a rule, students' contributions were short, included hardly any questions and were used as cue for a teacher contribution in up to 50% of all cases.

Summary and discussion. If the studies reviewed can be considered representative, one must conclude that classroom discourse in grammar lessons is typically restrained and cognitively undemanding, and that both students and teachers often stumble into perplexity. In students, problems manifest in the first instance as a failure to understand what is targeted by discourse on syntactic features; they also appear as struggles with formulations which need to be devised when applying grammatical operations. On the other hand, situations where understanding is established in a sudden and unexpected fashion seem to occur as well. In the studies under consideration, it is sometimes assumed that communication problems in grammar lessons result from students' inability to understand what the communication is about. However, the observation described above suggests an alternative explanation: students may, in principle, be able to grasp the syntactic features at stake yet be unable to stick to them in continuing dialogue when other

syntactic features interfere. The conclusion might be that the key problem of classroom discourse in grammar lessons is not how to gain access to syntactic information but rather how to keep track of it in dialogues which inevitably amass additional syntactic information.

What is meant by this may be illustrated by an incident reported by Bremerich-Vos (1996) (see above). In this incident, grade 5 children end up bedeviled when discussing 'whether you can see feeling' and 'whether you can touch running'. At first glance the incident seems to corroborate the widespread assumption that children attend to meaning instead of to linguistic form (see, e.g., Gornik, 1989) because students talk about the processes of feeling or running instead of talking about the words 'to feel' and 'to run'. However, that argument is not compelling. The children try to apply a criterion taught in primary schools to discern nouns, namely that nouns denote things you can see or touch. This criterion appears to be a semantic one but it actually works as a syntactic one if one takes it as saying that a noun may be inserted in a frame of the type

The/a/∅… you can see/touch

Depending on which noun one inserts, a true or a false statement results. If, however, one inserts a verb, one does not obtain a true or a false statement but an unintelligible sequence of words.

**Feeling you can see/*Running you can touch*

What the children discuss in the Bremerich-Vos example is one such unintelligible sequence of words. They seem to stray from the track because by trying to apply the criterion they are led to overstretch a syntactic template which normally guides their talk, and they cannot get back on track subsequently. Note that when one applies the above-mentioned noun criterion, two uses of the words 'feeling' and 'running' emerge: an object language use where the word is taken in its proper sense, and a metalanguage use where it is quoted. In the first use, the expressions 'feeling' and 'running' function as verbs, in the second use, they function as nouns. As a consequence, two syntactic templates, a verb and a noun template, compete and may interfere. The key problem, then, would be to maintain the intended template without being confused by the occurrence of the other one. Note that this problem is very similar to what Funke (2005) observed when using direct access tasks with grade 5-7 students.

These considerations concur with Wolt's (1974) assumption that the problems of discourse in grammar lessons are specifically connected to the topic of communication. In contrast to Wolt, however, they do not profile the difficulty as making language a distal object of attention but see it instead as maintaining proximal syntactic information even after first-pass linguistic processing.

4. GENERAL DISCUSSION

The use of grammar instruction. In the international debate, the 'big question' has been what use grammar instruction has as part of L1 education. What does the empirical research reviewed contribute to answer this question?

Doubts about the value of grammar instruction have mainly been based on two criticisms: First, it does not demonstrably foster language skills; second, it does not even build sustained grammatical knowledge (Hillocks & Smith, 2003).

As for the first question, no study has been conducted in German speaking regions so far to explore whether grammar instruction makes the overall quality of students' writing better. However, the results found in some Germanophone studies concur with international research in that grammar instruction may influence the language students use in writing on condition that it focuses on the functions of syntactic devices in the texts students write. Whether grammar instruction fosters reading comprehension has been examined in some Germanophone studies, but, just as in international studies, no evidence was found. Yet, some studies suggest that it might at least be worthwhile to further explore the impact of working on grammar on components of reading comprehension such as the grasp of information structure and the use of orthographic cues to syntactic structure. In sum, contributions of Germanophone research to the first question are sparse, but they are suited to corroborate and to amplify to some degree what has been found in international research.

As for the second question, the results found in a cohort of quasi-experimental Germanophone studies may be summarized as showing that students who were exposed to an intensified grammar instruction performed better on school-type tests of explicit grammar knowledge than students in regular classrooms. Also, their advantage was shown to be maintained over some weeks.

It is remarkable that these findings contrast with the impression aired by experts. For instance, Claus-Schulze (1966) assessed secondary students' explicit grammar knowledge as 'insufficient'. The linguist Habermann (2013) reported about huge gaps in the explicit grammar knowledge of undergraduate students. Following this, any grammatical knowledge which students might have when they are at school seems to evaporate by the time they enroll in university.

What should one conclude from such opposing assertions? They may be less incompatible than they seem at first sight. Though the learning outcomes in well-controlled intervention studies are modest, there are indications that students working on grammar at school may indeed gain some in-depth knowledge beyond simply memorizing facts which they subsequently forget. For instance, in five out of seven independent samples, Altendorf (1969), Friedrich (1970) and Siegert (1977) found that experimental students outperformed control students by making fewer erroneous classifications, and not by making more correct classifications. This is what one should expect if students learned about delimiting syntactical phenomena. The disastrous findings pertaining to adult university students which Habermann (2013) reports might mean that students forgot concepts of school grammar

and grammatical terminology. They do not, however, provide evidence that students lost sight of syntactic features as well; on the contrary, students' sensitivity to such syntactic features might be deeper than before. Recall that, as Riehme (1972) found, older students tended to formulate their knowledge about nouns in a sketchier, less terminological fashion than younger students, and that Funke (2005) reports some dissociation between students' school-type grammar knowledge and their achievements on direct access tasks. If so, a school-type test assessing terminological knowledge will not capture how such students access syntactic structures. Thus grammar instruction may be assumed to contribute to grammatical knowledge in students to some extent, and as long as one's definition of grammatical knowledge is not limited to explicit knowledge of school grammar. The really basic question with which one is left, then, is: What is grammatical knowledge if it is not simply explicit knowledge of grammatical concepts?

Active and efficient grammar knowledge. Two basic ideas about what active and efficient grammar knowledge actually is emerge from the studies reviewed. According to one idea, the key point is whether grammatical knowledge becomes available to someone independently of whether a specific retrieval context is given. This may be traced back to Vygotsky (1973, first published 1934), who assumed that knowledge is 'arbitrarily' accessible if its primary context of retrieval has been generalized. In other words, active and efficient grammatical knowledge is conceptual by nature. This idea, which might be called the epistemic view of grammatical knowledge, underlies most of the intervention studies, most notably those which aim at fostering students' conceptual grammatical knowledge. Another idea is that grammatical knowledge is active and efficient if it mirrors the structures sentences have. It becomes evident in the work of Döhmann (1977) who, based on generative grammar, assumes that the structure of a sentence is tied to the sentence as an inherent feature. This idea might be called the representational view of grammatical knowledge.

Review of the empirical studies suggests that neither the epistemic nor the representational view is suited to describing what active and efficient grammatical knowledge is. The main shortcomings of instructional approaches following the epistemic view result from the fact that they have students generalize features of linguistic form or meaning to arrive at grammatical concepts. However, focusing on form or meaning does not guarantee that one uses them properly as means of alerting oneself to the occurrence of syntactic features. Jantos (1971) found that children, after receiving instruction with a focus on form, attended to grammatical morphemes instead of syntactic configurations. Spies (1989) reports that students tended to equate syntactic categories with semantic features instead of syntactic roles. In sum, the epistemic view, when enlisted in instructional practice, replaces syntactic features with non-syntactic surrogates. This risks leading students astray, prompting them to focus on the latter instead of the former. The shortcoming of the representational view is that it has not been able to explain how students deal with syntactic assignments. Döhmann (1977) expected that students would spontaneously segment sentences into two main parts, a noun phrase part and a verb

phrase part, because the then-current generative model assumed sentences to be bi-partitioned in that way. However, this was not borne out in her data. In addition, an observation by Claus-Schulze (1966) suggests that how students classify the verbal parts of a sentence depends not on the sentence's structural representation but rather on how likely it is that syntactic information is prominent when the sentence is processed.

An alternative to the epistemic and representational views of grammar knowledge might be labeled an indexical view (cf. Pateman, 1987). It hypothesizes that a metasyntactic achievement related to a current utterance will be reliably accomplished if students are led by syntactic information which they generate when processing the utterance. The indexical view contrasts with the epistemic view by assuming that letting oneself be guided by self-generated syntactic information is a more powerful way to detect the syntactic structure of an utterance than looking at its linguistic form. It contrasts with the representational view by rejecting the idea that grasping an utterance's structure implies accessing a pre-specified representation. The indexical view suggests distinguishing between the access someone has to a syntactic feature and the description he or she is able to give of that feature. Having access to a syntactic feature does not imply the ability to describe it, and being able to describe a syntactic feature does not necessarily imply one has access to it.

From this it follows that in order to assess active and efficient grammar knowledge, one should not rely solely on data about students' explicit knowledge. Empirically tracing students' progress in accessing indexical syntactic information might reveal more improvement than is found when one uses explicit grammar tests. On the other hand, it might also reveal learning problems students experience which have gone unnoticed so far.

As stated in the Introduction section, the questions driving this review are conceptual as well as factual. From this, one might get the impression that the issues it deals with are metatheoretical and therefore hardly relevant to instructional practice. This is, however, not true. First, as demonstrated above, employing more appropriate methods of testing grammatical knowledge to evaluate grammar instruction might lead to a new picture of the knowledge students gain in such instruction, and hence a new understanding of the use of grammar instruction in general. Second, and more importantly, not only researchers but also teachers and even students have preconceptions about what people learn or should learn in grammar instruction. These may be assumed to deeply influence what happens in grammar classrooms. So, as a matter of fact, conceptual questions are eminently practical questions.

Limitations. Finally, the limitations of the present review should be elucidated. The empirical study of grammar teaching and learning in German classrooms is, on the whole, methodologically underdeveloped. As for quasi-experimental research, appropriate statistical treatment is found only in studies published after 2000. The existing microanalytic investigations rely on sophisticated interpretive procedures but analyze their data selectively. Several authors report amassing corpora com-

prising 10-20 videotaped grammar lessons each (Bremerich-Vos, 1996; Kleinbub, 2012; Spreckels, 2009); however, their published papers consider fewer than 15% of these.

To cope with these problems with respect to the quasi-experimental studies, effect sizes were reported instead of significance tests. This may compensate for inadequate statistical treatment, but one cannot be quite sure that the effects found are representative of the effects in general. In addition, the use of formulas which take into account the clustering of data (Hedges, 2009) prevents premature conclusions about effects, but also reduces the statistical power of homogeneity tests exerted on the effect sizes. It is possible that in some cases (e.g., the unspecific reading measures, Table 2) the effects are not homogeneous even though the statistic indicates homogeneity. As for the microanalytic studies, these studies were considered indicative of students' comprehension problems in accordance with the authors' views. However, one cannot rule out the possibility that lessons were selectively considered for analysis in published papers just because they displayed comprehension problems. As a result, one should consider that the microanalytic studies might function like concave mirrors: They show a real feature of grammar instruction, but they might artificially magnify that feature whilst downscaling or even distorting the rest of the picture.

5. CONCLUSION

Two tasks for future research emerge from the present review. First, it is necessary to develop tests of grammar knowledge which provide a direct, unambiguous focus on syntactic features without relying on a specific terminology or on indirect hints at which syntactic features are at stake. Tests which do not comply with this will, on a case-by-case-basis, over- or underestimate students' grammatical knowledge. Second, the main problem of grammar learning at school does not seem to be that students lack access to syntactic information but that they have difficulties maintaining continued access to it (Funke & Sieger, 2012). If so, L1 instruction must not presuppose that students enjoy continued access to syntactic information but must instead contribute to its emergence beforehand. Future research should explore whether and how this might be possible.

REFERENCES

Abraham, U. (2010). Grammatik beim Lesen und Schreiben von Texten in der Schule [Grammar in reading and writing texts at school]. In M. Habermann (ed.), *Grammatik wozu?* (pp. 323-340). Mannheim, Germany: Dudenverlag.

Altendorf, L. (1969). *Untersuchungen zur rückgreifenden Behandlung eines grammatischen Mittels im Prozeß der Erstvermittlung von Wissen unter besonderer Berücksichtigung des Vergleichs. Dargestellt an Stoffeinheiten der 5. Klasse* [Studies on the treatment of a grammatical device when teaching it for the first time with special regard to doing retrospective comparisons. Exemplified by teaching units in grade 5]. Dissertation Humboldt-Universität Berlin.

Andresen, H. (1996). 'Dann muß der Keller wohl eine Sache sein'. Thesen zum Grammatikunterricht und seinen Folgen [‚So the basement must be a thing'. On grammar instruction and its consequences]. *Grundschulunterricht, 43*(9), 9-12.

BIFIE (2006a). *Bildungsstandards für Deutsch 4. Schulstufe* [Educational standards for German studies, grade 4]. Salzburg, Austria: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens. https://www.bifie.at/node/1345

BIFIE (2006b). *Bildungsstandards für Deutsch 8. Schulstufe* [Educational standards for German studies, grade 8]. Salzburg, Austria: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens. https://www.bifie.at/node/325

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1-11.

Boettcher, W. (1994). Grammatikunterricht in Schule und Lehrerausbildung [Grammar instruction in school and in teacher education]. *Der Deutschunterricht*, *46*(5), 8-31.

Boettcher, W. (1999). Der Kampf mit dem Präpositionalobjekt [Struggling with the prepositional object]. In A. Bremerich-Vos (ed.), *Zur Praxis des Grammatikunterrichts* (pp. 193-252). Freiburg, Germany: Fillibach.

Bremerich-Vos, A. (1995). ‚Dann probiert mal schön!' – Mikroskopisches zur Bildung vom grammatischem Wissen im schulischen Unterricht ['So go ahead and try!' Microscopic views on the formation of grammatical knowledge in school-based instruction]. *Mitteilungen des Deutschen Germanistenverbandes, 42*(2), 27-32.

Bremerich-Vos, A. (1996). Deutschdidaktik und Unterrichtsforschung. Versuche in einem bislang vernachlässigten Feld [Didactics of German and research on instruction: An essay on a neglected field]. In G. W. Schnaitmann (ed.), *Theorie und Praxis der Unterrichtsforschung. Methodologische und praktische Ansätze zur Erforschung von Lernprozessen* (pp. 209-233). Donauwörth, Germany: Auer.

Bremerich-Vos, A. & Böhme, K. (2009). Kompetenzdiagnostik im Bereich ‚Sprache und Sprachgebrauch untersuchen' [Diagnosing competences within the domain of analyzing language and language use]. In D. Ganzer, O. Köller & A. Bremerich-Vos (eds.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (pp. 376-392). Weinheim, Germany: Beltz.

Brünner, G. (1982). 'Wer oder was kennst du?' Probleme des Grammatikunterrichts in der Grundschule ['Who or what do you know?' Problems of grammar instruction in the primary school]. In K. Detering, K. & W. Schmidt-Radefeld (eds.), *Sprache beschreiben und erklären*. Akten des 16. Linguistischen Kolloquiums Kiel 1981. Band 1 (pp. 136-146). Tübingen, Germany: Niemeyer.

Claus-Schulze, A. (1966). *Ergebnisse des Grammatikunterrichts in ihrer Beziehung zur Denkleistung der Schüler* [Results of grammar instruction and their relation to students' cognitive achievement]. Habilitationsschrift Pädagogische Hochschule Potsdam.

Claus-Schulze, A. (1978). Probleme und Ergebnisse von Untersuchungen zur Beziehung zwischen grammatischem Wissen und muttersprachlichem Können [Problems and results of research about the relationship between grammatical knowledge and verbal competence]. *Wissenschaftliche Zeitschrift der Pädagogischen Hochschule „Lieselotte Herrmann" Güstrow, 16*(1), 109-127.

Claus-Schulze, A. (1982). *Zur Rolle der Sprache im Unterricht und zur Methodologie der Erfassung von Sprachkönnen* [On the role of language in instruction and on the methodology of assessing language competence]. Berlin, Germany: Akademie der Pädagogischen Wissenschaften der DDR.

Delnoy, R., Herrrlitz, W., Kroon, S., & Sturm, J. (eds.) (1988). *Portraits in mother tongue education. Teacher diaries as a starting point for comparative research into standard language teaching in Europe*. Enschede, The Netherlands: IMEN.

Delnoy, R. & Kroon, S. (eds.) (1990). *Tisvildeleje Papers. A report of the IMEN-workshop, Tisvildeleje, Denmark, 26-29 May 1989*. Enschede, The Netherlands: IMEN.

Döhmann, U. (1977). Untersuchungen zum Grammatikunterricht [Investigations on grammar instruction]. Tübingen, Germany: Niemeyer. http://dx.doi.org/10.1515/9783111712680

EDK (2007). *Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat)* [Intercantonal agreement on the harmonisation of compulsory education]. Bern, Switzerland: Schweizerische Konferenz der kantonalen Erziehungsdirektoren. edudoc.ch/record/24711/files/HarmoS_d.pdf

Eichler, W. (1992). *Schreibenlernen* [Learning to write]. Bochum, Germany: Kamp.

Eichler, W (2007a). Sprachbewusstheit [Language awareness]. In B. Beck & E. Klieme (eds.), *Sprachliche Kompetenzen. Konzepte und Messung* (pp. 147-157). Weinheim, Germany: Beltz.

Eichler, W. (2007b). Sprachbewusstheit bei DESI [Language awareness in the DESI study]. In H. Willenberg (ed.), *Kompetenzhandbuch für den Deutschunterricht* (pp. 124-133). Baltmannsweiler, Germany: Schneider.

Elley, W., Barham, I., Lamb, H., & Wyllie, M. (1979). The role of grammar in a secondary school English curriculum. *Research in the Teaching of English, 10*(1) 5–21.

Erb, H. (1982). *Zur Bedeutung systematischer Selbstkontrolle sprachlicher Leistungen durch die Schüler für die Beherrschung sprachlicher Normen – dargestellt an der Beherrschung von Normen der Interpunktion in Klasse 5* [On the import of students' systematic self-control of linguistic achievements on the command of linguistic norms – as evidenced by students' command of punctuation norms in grade 5]. Dissertation Pädagogische Hochschule „Lieselotte Herrmann" Güstrow.

Friedrich, B. (1970). *Zur simultanen Erstvermittlung von Merkmalen zweier grammatischer Begriffe* [On the simultaneous teaching of two grammatical concepts]. Dissertation Humboldt-Universität Berlin.

Friedrich, B. (1996). Dissertationsthemen in der DDR [The topics of doctoral dissertations in the GDR]. *Didaktik Deutsch, 1*(1), 83-94.

Funke, R. (1995a). Das Heben des Wortartschatzes. Nomen im Kontext sehen [Yielding the word class hoard. Looking at nouns in context]. *Praxis Deutsch, 22*(129), 57-60.

Funke, R. (1995b). Grammatik in Funktion. Beobachtungen zu Groß- und Kleinschreibfehlern in Diktaten [Grammar in use. Observations on capitalization errors in written dictations]. *Deutschunterricht, 48*(9), 430-434.

Funke, R. (2005). *Sprachliches im Blickfeld des Wissens* [The explicitness of linguistic knowledge]. Tübingen, Germany: Niemeyer. http://dx.doi.org/10.1515/9783110924701

Funke, R. (2013). Syntaktisches Lernen im Schulalter und der Deutschunterricht [School-aged children's syntactic learning and German studies]. In T. Zybatow & U. Harendarski (eds.), *Sprechen, Denken und Empfinden* (pp. 157-172). Berlin, Germany: LIT Verlag.

Funke, R. (2014). Grammatikunterricht, grammatisches Wissen und schriftsprachliches Können [Grammar instruction, grammatical knowledge and literacy-related competences]. In H. Gornik (ed.), *Sprachreflexion und Grammatikunterricht* (pp. 429-456). Baltmannsweiler, Germany: Schneider.

Funke, R. & Sieger, J. (2012). Continued access to syntactic information in reading. *Reading and Writing: An Interdisciplinary Journal, 25*(7), 1769-1794. http://dx.doi.org/10.1007/s11145-011-9341-8

Funke, R., Wieland, R., Schönenberg, S. & Melzer, F. (2013). Exploring syntactic structures in first-language education: effects on literacy-related achievements. *L1-Educational Studies in Language and Literature, 13*, 1-24. http://dx.doi.org/10.17239/L1ESLL-2013.01.03

Galperin, P. J. (1972, first published in Russian 1957). Die geistige Handlung als Grundlage für die Bildung von Gedanken und Vorstellungen [Mental activity as the foundation for the formation of thoughts and ideas]. In P. J. Galperin & A. N. Leont'ev, *Probleme der Lerntheorie* (pp. 33-49). Berlin, Germany: Volk und Wissen 1972.

Glinz, H. (1963). Der Sprachunterricht im engeren Sinne oder Sprachlehre und Sprachkunde [Language instruction or system-oriented and phenomenon-related approaches to teaching first language]. In A. Beinlich (ed.), *Handbuch des Deutschunterrichts im ersten bis zehnten Schuljahr*. Erster Band. Third edition (pp. 225-323). Emsdetten, Germany: Lechte.

Gornik, H. (1989). Metasprachliche Entwicklung bei Kindern [Metalinguistic development in children]. In E. Haueis (ed.), *Sprachbewußtheit und Schulgrammatik* (pp. 39-58). Osnabrücker Beiträge zur Sprachtheorie, 40. Osnabrück, Germany: Verein zur Förderung der Sprachwissenschaft in Forschung und Ausbildung.

Graham, S. & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476.

Granzow-Emden, M. (1999). Grammatik ist, wenn man trotzdem fragt. Vom ‚Elend der Grammatiker-Fragen' [If you ask nevertheless, it's grammar. On the misery of grammarians' questions]. In P. Klotz & A. Peyer (ed.), *Wege und Irrwege sprachlich-grammatischer Sozialisation. Bestandsaufnahme–Reflexionen–Impulse* (pp. 169-184). Baltmannsweiler, Germany: Schneider.

Grimm, J. (1819). *Deutsche Grammatik. Erster Teil* [German grammar. Part I]. Reprint 1989. Hildesheim, Germany: Olms.

Habermann, M. (2013). Von der Schule zur Universität. Zum Funktionswandel von Grammatik im BA-Studium [From school to university. The functional change of grammar in undergraduate studies]. In K.-M. Köpcke & A. Ziegler (eds.), *Schulgrammatik und Sprachunterricht im Wandel* (pp. 35-60). Berlin, Germany: de Gruyter.

Harris, R. J. (1962). *An experimental inquiry into the functions and value of formal grammar in the teaching of English, with special reference to the teaching of correct written English to children aged twelve to fourteen*. PhD Thesis. London, UK: University of London.

Hedges, L. V. (2009). Effect sizes in nested designs. In H. Cooper, L. V. Hedges & J. C. Valentine (eds.), *The handbook of research synthesis and meta-analysis* (pp. 337-356). Second edition. New York, NY: Russell Sage Foundation.

Helmke, A. & Hosenfeld, I. (2007). *Ergebnisbericht VERA 2006/07* [Report on the VERA 2006-07 results]. Landau, Germany: Universität Koblenz-Landau. http://www.isq-bb.de/uploads/media/VERA_BE_Landesbericht.pdf

Heringer, H.-J. (2000). Wie die Grammatik beim Schreiben hilft [How grammar supports written formulation]. *Der Deutschunterricht, 52*(4), 21-28.

Heringer, H.-J. (2001). *Lesen lehren lernen. Eine rezeptive Grammatik des Deutschen* [Reading – teaching – learning. A receptive grammar of German]. Second edition. Tübingen, Germany: Niemeyer.

Herrlitz, W. (1990). Comparing mother tongue education in different European countries. In R. Delnoy & S. Kroon (eds.) (1990). *Tisvildeleje Papers. A report of the IMEN-workshop, Tisvildeleje, Denmark, 26-29 May 1989* (pp. 5-19). Enschede, The Netherlands: IMEN.

Herrmann, H. (1974). *Untersuchungen zur Effektivität der Begriffsbildung im Grammatikunterricht* [Studies of the efficiency of concept formation in grammar instruction]. Dissertation Pädagogische Hochschule Potsdam.

Hillocks, G. & Smith, M. W. (2003). Grammars and literacy learning. In J. Flood, D. Lapp, J. R. Squire & J. M. Jensen. (Eds.). *Handbook of research on teaching the English language arts*. 2nd edition (pp. 721-737). Mahwah, NJ: Erlbaum.

Hoffmann, L. (2000). Formulieren: Ein Fall für Grammatik [Formulating: A case for grammar]. *Der Deutschunterricht, 52*(4), 6-20.

Hohm, M. (2007). *Sprachbewusstheit, Lesekompetenz und Textverstehen* [Lingustic awareness, reading competence and text comprehension]. Saarbrücken, Germany: VDM.

Ingendahl, W. (1997). *Sprachreflexion statt Grammatik* [Reflection on language instead of grammar]. Tübingen, Germany: Niemeyer.

Isaac, K., Eichler, W. & Hosenfeld, I. (2008). Ein Modell zur Vorhersage von Aufgabenschwierigkeiten im Kompetenzbereich Sprache und Sprachgebrauch untersuchen [A model for the prediction of task difficulties within the domain of analyzing language and language use]. In B. Hofmann & R. Valtin (eds.), *Checkpoint Literacy. Tagungsband 2 zum 15. Europäischen Lesekongress 2007 in Berlin* (pp.12-27). Berlin, Germany: DGLS.

Isaac, K., Hochweber, J.& Eichler, W. (2010). Differenzierende Modellierung der Aufgabenschwierigkeit im Bereich ‚Sprache und Sprachgebrauch untersuchen' [Differentially modeling task difficulties within the domain of analyzing language and language use]. In B. Schwarz, P. Nenniger & R. Jäger (eds.), *Erziehungswissenschaftliche Forschung – nachhaltige Bildung. Beiträge zur 5. DGfE-Sektionstagung ‚Empirische Bildungsforschung'–AEPF-KBBB im Frühjahr 2009* (pp. 279-286). Landau, Germany: Verlag Empirische Pädagogik.

Isaac, K. & Hochweber, J. (2011). Modellierung von Kompetenzen im Bereich ‚Sprache und Sprachgebrauch untersuchen' mit schwierigkeitsbestimmenden Aufgabenmerkmalen [Modeling competences within the domain 'Study of language and language use' using task features which determine task difficulties]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *43*(4), 186-199. http://dx.doi.org/10.1026/0049-8637/a000054

Jantos, W. (1971). *Zur Entwicklung des Abstrahierens und Verallgemeinerns im Grammatikunterricht der 2. und 3. Klasse* [On the development of abstraction and generalization in grammar instruction in grades 2 and 3]. Dissertation Akademie der Pädagogischen Wissenschaften der DDR, Berlin.

Jantos, W. (1975). Die Entwicklung des Abstrahierens und Verallgemeinerns im Grammatikunterricht der Klassen 2 und 3 [The development of abstraction and generalization in grammar instruction in grades 2 and 3]. In J. Lompscher (ed.), *Theoretische und experimentelle Untersuchungen zur Entwicklung geistiger Fähigkeiten* (pp. 133-176). Second edition. Berlin, Germany: Volk und Wissen.

Keller, F. & Moser, U. (2012). *Schlussprüfung an der Volksschule des Kantons Basel-Stadt. Schlussbericht 2012* [Final examination in combined primary and lower secondary schools of the canton Basel City. Final report 2012]. Zürich, Switzerland: IfBE. http://www.ibc.uzh.ch

Kleinbub, I. (2012). Kontrastive Analyse von Zugängen zum Satzglied durch Unterrichtsvideos und -transkripte [Contrastive analysis of approaches to sentence parts based on videos and transcripts of school lessons]. In S. Merten & K. Kuhs (eds.), *Perspektiven empirischer Sprachdidaktik* (pp. 197-219). Trier, Germany: Wissenschaftlicher Verlag.

Kleinbub, I. (2014). Grammatik unterrichten – Professionsorientierung in der Lehrerbildung [Teaching grammar – Professional orientations in teacher formation]. In U. Bredel & C. Schmellentin (eds.), *Welche Grammatik braucht der Grammatikunterricht?* (pp. 135-160). Baltmannsweiler, Germany: Schneider.

Klotz, P. (1996). *Grammatische Wege zur Textgestaltungskompetenz. Theorie und Empirie* [Grammatical routes to composing text]. Tübingen, Germany: Niemeyer. http://dx.doi.org/10.1515/9783110940053

KMK (2004). *Bildungsstandards im Fach Deutsch für den Mittleren Bildungsabschluss* [Standards in German studies for the intermediate school leavingl qualification]. Herausgegeben vom Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Köln, Germany: Wolters Kluwer.

KMK (2005). *Bildungsstandards im Fach Deutsch für den Primarbereich* [Standards in German studies for primary education]. Herausgegeben vom Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Köln, Germany: Wolters Kluwer.

KMK (2014). *Kompetenzstufenmodell zu den Bildungsstandards im Kompetenzbereich ‚Sprache und Sprachgebrauch untersuchen' für den Mittleren Bildungsabschluss*. (Vorbehaltlich redaktioneller Änderungen) [Modeling competence levels with respect to the educational standards for the intermediate school leaving qualification pertaining to the domain of analyzing language and language use]. Berlin, Germany: IQB. http://www.iqb.hu-berlin.de/bista/ksm

Kochan, D. C. (1975). *Forschungen zum Deutschunterricht. Ergänzte Teilausgabe des Handbuchs zur Unterrichtsforschung* [Research on German studies. Supplemented partial edition of the handbook of instructional research]. Weinheim, Germany: Beltz.

König, S. (1992). *Zur Begriffsbildung bei der Bewältigung von Anforderungen des Grammatikunterrichts der Klassen 5–7 am Beispiel der Wortarten* [Formation of concepts when dealing with the challenges of grammar instruction in grades 5-7, exemplified by reference to word classes]. Dissertation Pädagogische Hochschule Magdeburg.

Konsortium HarmoS Schulsprache (2010). *Schulsprache. Wissenschaftlicher Kurzbericht und Kompetenzmodell* [Language of the school. Technical summary report and competence model]. http://www.edk.ch/dyn/11659.php

Kramarczyk, A. & Wazel, G. (1968). *Lernprogramme im Grammatikunterricht* [Progammed learning in grammar instruction]. Dissertation Friedrich-Schiller-Universität Jena.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge, UK. Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511620607

Lehmann, R. H., Peek, R. & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern, die im Schuljahr 1996/97 eine fünfte Klasse an Hamburger Schulen besuchten. Bericht über die Erhebung im September 1996 (LAU 5)* [Starting conditions and learning development in students enrolled in a grade 5 class in Hamburg schools during the 1996-97 school year. Report on the survey in September 1996 (LAU 5)]. Hamburg, Germany: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule. http://bildungsserver.hamburg.de/contentblob/2815702/data/pdf-schulleistungstest-lau-5.pdf

Lehmann, R. H., Peek, R. & Gänsfuß, R. (1999). *Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern, die im Schuljahr 1996/97 eine fünfte Klasse an Hamburger Schulen besuchten. Bericht über die Erhebung im September 1998 (LAU 7)* [Starting conditions and learning development in students enrolled in a grade 5 class in Hamburg schools during the 1996-97 school year. Report on the survey in September 1998 (LAU 7)]. Hamburg: Germany: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule. http://bildungsserver.hamburg.de/content-blob/2815702/data/pdf-schulleistungstest-lau-7.pdf

Lehmann, R. H., Peek, R., Gänsfuß, R., & Husfeldt, V. (n. d.). *LAU 9. Aspekte der Lernausgangslage und der Lernentwicklung Klassenstufe 9. Ergebnisse einer längsschnittlichen Untersuchung in Hamburg* [LAU 9. Starting conditions and learning development in grade 9. Results of a longitudinal survey in Hamburg]. http://www2.bildungsserver.hamburg.de/contentblob/2815692/data/pdf-schulleistungstest-lau-9.pdf

Mehan, H. (1994). The role of discourse in learning, schooling, and reform. In B. McLeod (ed.), *Language and learning. Educating linguistically diverse students* (pp. 71-96). Albany, NY: State University of New York Press.

Melzer, F. (2011). Direktes Umgehen mit syntaktischen Strukturen: ein Unterrichtsansatz für die Sekundarstufe I [Direct exploration of syntactic structures: An instructional approach for secondary school classes]. In C. Noack & J. Ossner (eds.), *Grammatikunterricht und Grammatikterminologie. Osnabrücker Beiträge zur Sprachtheorie 79* (pp. 159-180). Duisburg, Germany: Universitätsverlag Rhein-Ruhr.

Mesch, B. & Dammert, Y. (2015). Verbwissen in der Primarstufe [Verb knowledge in the primary grades]. In B. Mesch & B. Rothstein (eds.), *Was tun mit dem Verb?* (pp. 1-44). Berlin, Germany: de Gruyter.

Metz, K. (2005). *Grammatikkenntnisse – Kommasetzung* [Grammar knowledge - Punctuation]. Baltmannsweiler, Germany: Schneider.

Meyer, G. (1969). *Zur Begriffsbildung im Grammatikunterricht der Klasse 5 auf der Grundlage von Ergebnissen und Methoden der neueren Grammatikforschung* [On concept formation in grammar instruction based on the results and methods of recent grammar research in grade 5]. Dissertation Karl-Marx-Universität Leipzig.

Moser, U., Buff, A., Angelone, D., & Hollenweger, J. (2011). *Nach sechs Jahren Primarschule. Deutsch, Mathematik und motivational-emotionales Befinden am Ende der 6. Klasse*. Zürich, Switzerland: Bildungsdirektion des Kantons Zürich [After six years of primary school. German studies, mathematics, and socio-emotional condition at the end of grade 6]. Zürich, Switzerland: Bildungsdirektion Kanton Zürich. http://www.bi.zh.ch

Muhr, R. (1995). Ein schwieriges Verhältnis. Schulgrammatik versus Lernergrammatik im Deutschunterricht an Österreichs Schulen [A complicated relationship: School grammar versus learners' grammar in German studies in Austrian schools]. *Informationen zur Deutschdidaktik, 19*(3), 37-49.

Müller, A.& Tophinke, D. (2015). Verben als Lerngegenstand in der Sekundarstufe I: Was wissen Schülerinnen und Schüler – und was können sie wissen? [Verbs as a topic of instruction in the secondary grades: What knowledge do students have – and what knowledge may be accessible to them?] In B. Mesch & B. Rothstein (eds.), *Was tun mit dem Verb?* (pp. 46-78). Berlin, Germany: de Gruyter.

Myhill, D. A., Jones, S. M., Lines, H. & Watson, A. (2012). Re-thinking grammar: the impact of embedded grammar teaching on students' writing and students' metalinguistic understanding. *Research Papers in Education*, 27(2), 139-166

Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (eds.), *Vocabulary acquisition* (pp. 52-77). New York, NY: Guilford.

Pateman, T. (1987). *Language in mind and society. Studies in linguistic reproduction*. Oxford, UK: Clarendon Press.

Perfetti, C. A., Landi, N., & Oakhill, J. V. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (eds.), *The science of reading. A handbook* (pp. 227-247). Malden, MA: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch13

Peyer, A. (1997). Sätze [Sentences]. *Praxis Deutsch, 25*(147), 12-22.

Ploog, G. (1981). *Die Verbbehandlung in Klasse 5 unter besonderer Berücksichtigung der Gestaltung des Verhältnisses von Einheitlichkeit und Differenzierung beim Einsatz von Arbeitsverfahren* [Dealing with verbs in grade 5 by taking into special account the relationship between unification and differentation when using operative procedures]. Dissertation Humboldt-Universität Berlin.

Riehme, J. (1972). Zur Behandlung der Groß- und Kleinschreibung [Teaching capitalization]. *Deutschunterricht, 25*(9), 610-622.

Roeder, P. M. & Schümer, G. (1976). Unterricht als Sprachlernsituation [Schooling as an opportunity for language learning]. Düsseldorf, Germany: Schwann.

Rogers, L. A. & Graham, S. (2008). A meta-analysis of single-subject design writing intervention research. *Journal of Educational Psychology, 100*(4), 879–906.

Rose, K. (1997). Ist das nun ein Adjektiv, ein Substantiv oder was? Wie Schüler sprachliche Kategorienbegriffe repräsentieren - Versuch einer Annäherung [Is this an adjective, a noun or what? Steps toward understanding how students represent linguistic categories]. *Deutschunterricht, 50*(6), 394-401.

Sánchez-Meca, J., Marín-Martínez, F. & Chacón-Moscoso, S. (2003). Effects-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods* 8, 448-467. http://dx.doi.org/10.1037/1082-989X.8.4.448

Sappok, C. (2011). *Das deutsche Komma im Spiegel von Sprachdidaktik und Prosodieforschung* [The comma in German in light of the study of language learning and of prosody]. Berlin, Germany: LIT Verlag.

Scholz, H. (1975). *Untersuchungen zur Entwicklung von Können und Überzeugungen bei der Arbeit am grammatischen Mittel Attribut in den Klassen 5 und 7* [Studies in the development of competences and beliefs during instruction on the grammatical device of attribute in grades 5 and 7]. Dissertation (A). Leipzig: Karl-Marx-Universität Leipzig.

Schößler, H. (1981). *Die Ausbildung der Lerntätigkeit im vierten Schuljahr nach der Lehrstrategie des Aufsteigens vom Abstrakten zum Konkreten im Muttersprachunterricht (Bereich Syntax)* [The development of learning activity in grade 4 according to the method of ascending from abstract to concrete in first language instruction (domain of syntax)]. Dissertation Akademie der Pädagogischen Wissenschaften der DDR, Berlin.

Schübel, A. (2000). Lernstandserhebung im Fach Deutsch in den Klassenstufen 5 und 6. Teil II: Grammatisches Wissen und Können [A survey on learning outcomes in German studies in grades 5 and 6. Part II: Grammatical knowledge and competence]. *Deutschunterricht, 53*(6), 449-457.

Schumer, H. (1988). *Zur Herausbildung von Erkenntnisinteresse im Grammatikunterricht der Klassen 2 und 3. Dargestellt am Beispiel der Behandlung der Stoffeinheit 'Das Adjektiv'* [On the development of topic-related interest in grammar instruction in grades 2 and 3. Exemplified by the treatment of adjectives]. Dissertation Pädagogische Hochschule Erfurt.

Schuttkowski, C., Rothstein, B., Schmitz, A., & Gräsel, C. (2015). Lautes Denken in grammatikdidaktischen Fragestellungen? Diskussion zweier Studien [Thinking aloud as a method of research on teaching grammar? A discussion of two studies]. *Zeitschrift für Angewandte Linguistik* 63(1), 222-238. http://dx.doi.org/10.1515/zfal-2015-0016

Siebenbrodt, M. (1966). *Ein Beitrag zur Einwirkung des Grammatikunterrichts auf schriftliche Schüleräußerungen (dargestellt an Untersuchungen des Grammatikunterrichts vornehmlich 6. Klassen der zehnklassigen polytechnischen Oberschule)* [A contribution pertaining to the impact of grammar instruction on students‘ written utterances]. Habilitationsschrift Martin-Luther-Universität Halle.

Siegert, K. (1977). *Zum Einsatz von Identifikationsverfahren im Grammatikunterricht der allgemeinbildenden zehnklassigen Polytechnischen Oberschule. Untersuchungen zur Verbesserung des Muttersprachunterrichts* [The use of identification procedures in grammar instruction in polytechnical secondary schools. Studies on the improvement of mother tongue education]. Dissertation Humboldt-Universität Berlin.

Sijtstra, J., van der Schoot, F. & Hemker, B. (2002). *Balans van het taalonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 1998* [Record of first language instruction at the end of grade 3 of the primary school]. Arnhem: Citogroep.

Simmel, C. (2007). Wie erklären sich Schülerinnen und Schüler gegenseitig grammatische Phänomene? [How do students explain grammatical phenomena to each other?]. In E. Haueis & S. Schallenberger (eds.). *(Schrift-)Spracherwerb und Grammati(kali)sierung Osnabrücker Beiträge zur Sprachtheorie 73* (pp. 57-80). Duisburg, Germany: Gilles & Francke.

Spies, B. (1989). Zur Aneignung von grammatischem Wissen bei Primarstufenschülern [On the acquisition of grammatical knowledge in primary grades students]. In E. Haueis (ed.), *Sprachbewußtheit und Schulgrammatik* (pp. 75-85). Osnabrücker Beiträge zur Sprachtheorie, 40. Osnabrück, Germany: Verein zur Förderung der Sprachwissenschaft in Forschung und Ausbildung.

Spreckels, J. (2009). Mündliches Erklären im Deutschunterricht [Oral explanations in German lessons]. In M. Krelle & C. Spiegel (eds.), *Sprechen und Kommunizieren* (pp. 117-138). Baltmannsweiler, Germany: Schneider.

Stahns, R. (2013). *Kognitive Aktivierung im Grammatikunterricht* [Cognitive activation in grammar instruction]. Baltmannsweiler, Germany: Schneider.

Stahns, R. (2014). Der Beitrag der empirischen Unterrichtsforschung zur Debatte um Unterrichtsqualität im Grammatikunterricht [The contribution of empirical research on instruction to the debate on instructional quality in grammar lessons]. In H. Gornik (ed.), *Sprachreflexion und Grammatikunterricht* (pp. 415-428). Baltmannsweiler, Germany: Schneider.

Studer, C. & Berger, S. (2013). *Check5 2012. Zwischenbericht zuhanden des Departements Bildung, Kultur und Sport des Kantons Aargau* [Check5 2012. Interim report of the Department for Education, Culture and Sports of the canton Aargau]. Zürich, Switzerland: IfBe. http://www.ibe.uzh.ch/projekte/check5.html

Symann, J. (1980). *Die Einheit des Muttersprachunterrichts als eine Voraussetzung zur rationellen Ausbildung des muttersprachlichen Könnens – dargestellt am Beispiel des Grammatikunterrichts in Klasse 5* [The unity of first language education as a precondition for efficient development of first language competence as exemplified by grammar instruction in grade 5]. Dissertation Akademie der Pädagogischen Wissenschaften der DDR, Berlin.

Szubert, S. (2015). Das „Anagramm-Training“. Ein exploratives Forschungsprojekt zur Verbesserung des sprachlichen Könnens ['Anagram training'. An explorative research study on the improvement of linguistic competence]. *Leseräume, 2*(2), 212-226. http://www.leseräume.de

Valtin, R. (1988). ‚<Gliedmaßen> ist ein Tu-Wort: Sie können weh tun' ['<Extremities> is a verb: They can ache']. *Grundschule, 20*(10), 26-27.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. http://www.jstatsoft.org/v36/i03/. http://dx.doi.org/10.18637/jss.v036.i03

Vygotsky, L. S. (1973, first published in Russian 1934). Denken und Sprechen [Thinking and speaking]. Frankfurt a. M, Germany: Fischer.

Weaver, P. A. (1979). Improving reading comprehension. Effects of sentence organization instruction. *Reading Research Quarterly, 15*(1), 129-146. http://dx.doi.org/10.2307/747435

Weingarten, R. (2001). Orthographisch-grammatisches Wissen [Orthographic-grammatical knowledge]. In S. Wichter & G. Antos (ed.), *Wissenstransfer zwischen Experten und Laien* (pp. 209-226). Frankurt a. M., Germany: Peter Lang.

Weisgerber, L. (1950). Grammatik im Kreuzfeuer [Grammar under crossfire]. *Wirkendes Wort, 1*(1), 129-139.

Wolt, R. (1974). *Untersuchungen zur Entwicklung des sprachlich-kommunikativen Könnens der Schüler als Prinzip des Unterrichts – dargestellt am Beispiel des Faches Deutsch* [Studies on the development of students' verbal and communicative competence as a guideline for instruction, as exemplified in German studies]. Dissertation Akademie der Pädagogischen Wissenschaften der DDR, Berlin.