Progress in Science Education, 2018
ISSN 2405-6057
DOI.10.25321/prise.2018.756

TEACHER WAIT-TIMES IN PHYSICS INSTRUCTION OF DIFFERENT DURATION (45 VS. 60/90 MINUTES)

Rainer Wackermann[a*], Philip Timmerman[a], Heiko Krabbe[a]
[a]Ruhr-University Bochum, Faculty of Physics and Astronomy, Physics Education Work-group, Universitätsstr. 150, 44801 Bochum, Germany
*Address all correspondence to Rainer Wackermann, wackermann@physik.rub.de

ABSTRACT
In recent times, some schools in Germany change the traditional 45-minutes lesson duration to 60 or 90 minutes. Little is known about the impact on the quality of instruction. Teachers' wait-time after asking questions is related to the cognitive activation of pupils. This exploratory study pursues the question whether a prolongation of lesson duration changes the average wait-time of four individual physics teachers. The finding is that the teachers under investigation show slightly longer wait-times than other teachers reported in literature, and that wait-time improvement does not happen automatically with lesson prolongation. In conclusion, teacher wait-time seems more likely to be a stable characteristic of the individual teacher. Wait-time improvement seemingly does not happen automatically with lesson prolongation, but probably needs sensitization and additional training.

Background: In recent times, some schools in Germany change the traditional 45-minutes lesson duration to 60 or 90 minutes. Little is known about the impact on the quality of instruction. Teachers' wait-time after asking questions is related to the cognitive activation of pupils.

Purpose: This exploratory study pursues the question whether a prolongation of lesson duration (45 vs. 60/90 minutes) changes the average wait-time of four individual physics teachers. It is the aim to formulate an evidence-based hypothesis whether teacher wait-times are a rather stable characteristic of the individual teacher, or dependent on the lesson duration.

Sample/setting: The participants comprise of four experienced physics teachers (one female, three male, from two different upper level secondary schools), who all in 2005/06 took part in a learning-process oriented teacher training and were then videotaped several times. Because their schools changed the lesson duration in 2009, one school to 60 minutes, the other to 90 minutes, it was now possible to videotape "longer" physics lessons of the same teachers. Total sample size is n=24 videos.

Design and methods: A pre-post comparison of the instruction of all four physics teachers was carried out using three short and three longer lessons per teacher. All whole-class dialogue was transcribed. Wait-times W1 and W2 were directly measured from the transcripts of the videos. Additional control variables shall give an estimate of the equality of corresponding short and long lessons. Further findings by Wackermann and Hater (2016) will be used to enrich a case-wise discussion.

Results: Although longer lessons should in principle allow for longer wait-times, the longer lessons under investigation here do show significantly varied wait-times with large effect sizes only for one teacher. In addition, those variations are not substantial enough to reach optimal wait times – and go both ways. However, the control variables shed some light on the individual teacher's use of the extra lesson time. A case-wise (teacher-wise) exploratory discussion shows that for three of the four teachers lesson prolongation leads to a somewhat different lesson concept, and that one teacher uses the extra time by adding new learning phases.

Conclusions/Implications for classroom practice and future research: In conclusion, teacher wait-time seems more likely to be a stable characteristic of the individual teacher. Wait-time improvement seemingly does not happen automatically with lesson prolongation, but probably needs sensitization and additional training. Further conclusions question the use of questions and the sensefulness of lesson prolongation to 90 minutes.

Keywords: Lesson duration, Quality of instruction, Question level, Stability of teacher characteristic, Video study, Wait-time

Accepted: August 2018

1. INTRODUCTION

In recent times, some schools in Germany change the traditional 45-minutes lesson duration to 60 or 90 minutes in an attempt to relieve schooling from too much unrest. This change happens basically without any empirical or theoretical justification, see Stender, Geller, Neumann & Fischer (2013) or Wackermann and Hater (2016). Little is known about the impact on the quality of instruction. It is well known, however, that teachers' wait-time after asking questions is related to the cognitive activation of pupils (Rowe, 1974; Tobin, 1987). Optimally long wait-times, see further below, can thus be regarded as an element of quality of instruction.

This exploratory study pursues the question whether a prolongation of lesson duration from 45 to 60 or 90 minutes changes the average wait-time of a sample of 24 lessons of four experienced German physics teachers.

2. RESEARCH BACKGROUND

Characteristic for German physics instruction is a content-developing classroom dialogue accompanied by teacher demonstrations (Labudde & Duit, 2007; Seidel, Prenzel, Rimmele, Dalehefte, Herweg, Kobarg & Schwindt, 2006). The whole-class dialogue is similar to a Socratic dialogue. By asking questions, teachers guide the instruction and set the pace of the interaction. Research in mathematics instruction in Germany (where physics teachers often also teach mathematics) shows that teachers ask many more questions than pupils and occupy dominantly more classroom dialogue time than pupils (Begehr, 2005), e.g. with a ratio of 3 to 1 in classroom dialogue time. Similar contemporary results from Germany come from other school subjects like English (DESI-Konsortium, 2006). The teacher dominance can be understood as rivalry about precious classroom time, although questions by pupils and the ability to express themselves are regarded as important for learning (Ackermann, 2011). There is ongoing related educational research concerning teacher questioning and classroom discourse in general (e.g. Hattie, 2009, p. 182) or in the science classrooms in particular (e.g. Eliasson, Karlsson & Sörensen, 2017), and a long history thereof (e.g. Corey, 1940).

One characteristic element of the interaction between teachers and their pupils is the so-called wait-time of teachers. In this study, teacher wait-time shall be used as a prerequisite to assess the cognitive activation of the pupils within whole-class dialogue.

2.1 Research on teacher wait-time

Already in the 1960s, M. B. Rowe, pioneer of research on wait-time and on the interaction of teachers and their pupils, defined two critical wait-times, independent of culture and educational level and still used in current research (Helmke, 2009; Ingram & Elliot, 2016; Leitz, 2015; Li & Arshad, 2014; Lipowsky, Rakoczy, Pauli, Reusser & Klieme, 2007).

(i) The „wait-time 1" (W1) is defined by the time that passes by after a teacher has asked a question until a pupil gives an answer.

(ii) The „wait-time 2" (W2) is defined by the time that passes by between an answer given by a pupil and the next question or expression of the teacher.

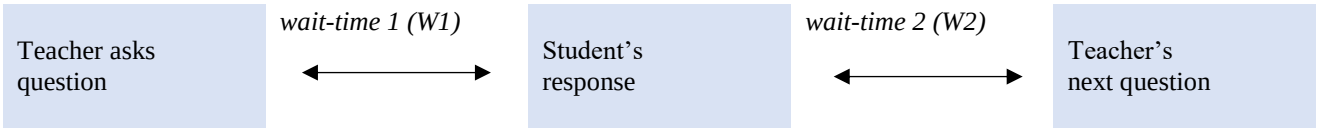

Fig. 1. Visualization of wait-times 1 and 2

The average wait-time of teachers, who are not sensitized in this respect, is roughly one second for both wait-times W1 and W2 (Rowe, 1974).

Important finding of research on wait-time is that there are supposedly optimal or recommended wait-times of in general more than three seconds (Helmke, 2009; Lipowsky et al., 2007; Leitz, 2015; Rowe, 1974; Tobin, 1987). Table 1 gives an overview of such optimal/recommended wait-times. With respect to wait-time 1 (W1), research has shown that it is important to further differentiate between low-level and high-level questions, where level is understood according to Bloom's taxonomy (1974). The difference between recommended optimal wait-times of up to 15 seconds and observed average wait-times for not-sensitized teachers of roughly one second seems substantial.

Tab. 1. Optimal wait-times.

	W1 – Low-level question	**W1 – High-level question**	**W2**
Optimal wait-time	3–4 seconds	Up to 15 seconds	More than 3 seconds

When wait-times are optimal, there are several beneficial consequences reported (Rowe, 1974; Tobin, 1987): Pupils give longer and more elaborate answers, are triggered to higher-level thinking, refer more to each other, and more individual pupils take part in a whole-class dialogue. In consequence, the performance of the pupils gets improved, although according to Tobin (1987) findings concerning the connection between wait-times and pupil performance are somewhat inconsistent.To sum up, the relatively easy measurement of wait-times allows conclusions about the cognitive activation of pupils within a whole-class dialogue.

2.2 Changes in physics classroom instruction due to lesson prolongation

For German physics instruction, lesson prolongation has lately been a topic of research (Stender et al., 2013; Zander, Krabbe, & Fischer, 2014; Wackermann & Hater, 2016; Hausen, Wackermann, & Krabbe, 2016). For instance, Zander et al. (2014) report that learning processes get completed only in double-lessons of 90 minutes and when teachers simultaneously undergo specific coaching. Wackermann and Hater (2016) as well as Hausen et al. (2016) report that longer lessons of 60 or 90 minutes show changes in the lesson structure, a higher educational variety has the potential for quality improvement for instance with respect to pupils' lab work and completion of learning processes. The findings allow the conclusion that further deep changes of classroom instruction due to lesson prolongation such as changes of wait-time are thinkable. But does a lesson prolongation really lead to longer wait-times potentially featuring more high-level questions? Or is teacher wait-time a stable characteristic of the individual teacher like other elements of the teacher-pupil interaction (Seidel & Prenzel, 2006)?

Thus, we formulate the following research question:

Does a prolongation of (German) physics lessons increase the average teacher wait-times W1 and W2?

It is our aim to formulate an evidence-based hypothesis whether teacher wait-times are a rather stable characteristic of the individual teacher, or dependent on the lesson duration.

3. METHODS

To answer the research question of stability of wait-times, it is necessary to investigate the instruction of the same teachers in a pre-post setting with varying lesson durations. Since lesson duration is a school-wide decision, which cannot be varied back and forth for individual teachers, it is nearly impossible to create experimental conditions with larger sample sizes. For this exploratory study, however, we were able to use the video data of two other small case studies that by chance offer suitable conditions.

The sample consists of four experienced physics teachers (one female, three male, from two different upper level secondary schools), who all in 2005/06 took part in a learning-process oriented teacher training (Wackermann, Trendel & Fischer, 2010). During the course of the teacher training, the teachers and one of their classes, 9th grade, were video-taped several times under the condition of 45-minute lessons. Both schools changed the lesson duration in 2009, one school to 60 minutes, the other one to double-lessons (90 minutes). Later, the four teachers with one of their classes, 9th grade again, were videotaped several times (2010/11 and 2013/14, respectively) under the condition of longer lessons. The lesson topics were not controlled, but covered ordinary 9th grade topics like mechanics and electricity.

The sample has already been analyzed by Wackermann and Hater (2016) and by Hausen et al. (2016). The four teachers were interviewed concerning their use of the extra lesson time by means of structured interviews (Wackermann & Hater, 2016; Hausen et al., 2016), videos were analyzed with respect to learning process orientation (Wackermann et al., 2010), and the pupils filled out questionnaires concerning their ability to follow the instruction (Wackermann et al. 2010; Wackermann & Hater, 2016). We will use results of these case studies to enrich our discussion later.

For this study, of each teacher three lessons of the pre and three lessons of the post recording sessions were taken – for one teacher, this is all the videos that are available, for the other teachers three videos were chosen as typical for the teacher and with extensive whole-class dialogue. So, total sample size is n=24 videos. We make the assumption that lessons from different years and with different classes, but from the same teachers, are comparable with respect to the average teachers' wait-time. Wait-time was never mentioned to the teachers, and their instruction is generally and under both conditions characterized by good classroom-management and a pupil-oriented attitude.

Independent variable in this study is the lesson duration (45 vs. 60/90 minutes). Dependent variables are *teacher wait-times W1 and W2* according to Rowe (1974).

Control variables are the *ratio of teacher talk time to pupil talk time* within a lesson and the *time share (percentage) of whole-class dialogue* within a lesson. The former gives an estimate of teacher dominance within whole-class dialogue, the latter gives an estimate of

equality of the structure of corresponding short and long lessons.

Additionally, the *question-level* gets estimated directly. We would expect teachers to maybe ask more high-level questions in longer lessons. Recently proposed systems (Heinze & Erhard, 2006; Patrick & Urhievwejire, 2012) draw on Bloom's taxonomy (1974), where the question level is differentiated between knowledge (identified by action verbs like name, list, describe, …), comprehension (explain, summarize, ...), application (apply, solve, ...), analysis (differentiate, point out, …), synthesis (predict, create, …) and evaluation (assess, judge, …). For the purpose of this study and as a simple first category system, we differentiate only between high- and low-level questions plus a category "other" e.g for classroom management purposes. Therefore, the higher five Bloom categories are all combined into high-level questions, and the knowledge level solely is kept as low-level questions. This simple differentiation is in accordance with long-time research (e.g. Corey, 1940; Haynes, 1935).

In view of the small sample size (four different teachers), this study is an exploratory case study with the aim of setting up evidence-based hypotheses concerning the stability of wait-times for lessons of different duration.

All whole-class dialogue was transcribed, excluded were class times of pupil lab work and of parallel group work as well as private side talk among pupils. Included were all teacher questions and expressions, naturally, and all pupil questions and expressions, regardless whether those were content-focused or not and regardless whether the pupils actually put their hands up or not.

The dependent variables teacher wait-times W1 and W2 are measured directly from the video according to their definitions. Table 2 shows an exemplary excerpt from a transcript, and the respective calculation of wait-times W1 and W2.

Tab. 2. Excerpt from a transcript.

Run time (mm:ss, start)	Run time (mm:ss, end)	Duration (mm:ss)	Who	Content/discourse
14:40	14:45	00:05	Teacher	„Is this rather a deformation or rather a change in motion? In effect?"
Wait-time 1		*00:01*		
14:46	14:47	00:01	Pupil	„Deformation."
Wait-time 2		*00:01*		
14:48	15:48	01:00	Teacher	„All agree? – Deformation. Right. Yes, this would be one possibility …"

Teacher wait-time W1 starts with the end of the (last) question, marked by the (last) question mark ("?") in the column "content/discourse". Wait-time W1 ends when the teacher allows the first answer or expression of a pupil (see fig. 1). The verbal or gestural call-up of the pupil is not considered and falls into wait-time W1. In the case shown in table 2 the measurement of W1 begins after "In effect?" and ends with the beginning of "Deformation.". An intercoder-reliability check for three of the 24 videos using two different raters concerning the set of time marks was highly satisfying with a 94% agreement (simple computation using # of agreed time marks/all time marks). Time marks are set with a precision of one full second. The remaining differences were in all instances one second, and mostly due to time marks positioned in between full seconds.

The average wait-times are simply computed for an entire lesson using the whole transcript of one lesson. Likewise, the other variables (density of teacher questions/level of questions, ratio of teacher vs. pupil questions/dominance and time share of whole-class dialogue within a lesson/structure) are simply counted from the entire transcript of one lesson.

The question level here gets assessed to be low-level, because it is a simple recall of information (Bloom level 1) and does not require higher-level thinking (Bloom levels 2–6). All teacher questions of all 24 videos were rated according to this scheme. An interrater-reliability check for three short lessons of three different teachers using the same two different raters as above is acceptable with a 83% agreement (simple computation using # of agreed ratings/all questions).

The planned analyses consist of description and of standard U-tests for check of significance and for computation of effect sizes (Cohen's d). The analyses will be carried out for each teacher individually and for each lesson duration separately, according to the study design.

4. RESULTS

We will first present findings concerning the control variables, see table 3, then findings concerning the dependent variables, see table 4.

For better clarity, table 3 displays averages of all three videos for each teacher and for each lesson duration. The full data set with each lesson individually can

be found in the appendix in table 3a. According to table 3, teachers ask many more questions than pupils, and teacher talk time typically outnumbers pupil talk time at least by a factor of 2, often 3, and up to 6.6. So the teachers tend to be quite dominant, which is in agreement with the research background (see earlier). Only in the case of teacher H under the condition of 90 minutes, the ratio of teacher to pupil talk time is balanced, but the ratio was 3.4 beforehand. For the other teachers, there is not much difference concerning teacher *dominance* between short and long lessons.

Absolute whole-class dialogue time (the sum of teacher and pupil talk time) lasts longer than 17 minutes for every lesson under investigation, in many lessons 30, 40 or more minutes. The time share (percentage) of whole-class dialogue within a lesson, in this study used as an estimate of equality of the *structure* of short and long lessons, ranges from 41 to 75 percent. For both 45 to 90 min teachers the time share (percentage) of whole-class dialogue decreases somewhat from short to long lesson duration, for the 45 to 60 min teachers it once increases and once stays nearly the same. Please note that the actual lesson duration in all cases differs somewhat from the nominal lesson duration (45/60/90 min), the difference being more pronounced for both 45 to 90 min teachers.

As further information, the *share of high-level questions* as a first indicator of the level of questions is given. It varies between 41% and 57%. It is especially high for teacher S under the condition of 45 minutes. It seems to rise somewhat for all other three teachers from short to long lessons approaching teacher S's question level. The amount of organizational questions, not displayed, lies around ten per lesson regardless of teacher or lesson duration with the exception of teacher S, who only poses on average four such managerial questions per lesson. So indeed around half of the teacher questions are high-level according to the definition used here (Bloom levels 2–6).

Tab. 3. Control variables: Number of questions by teacher and by pupils, absolute time of teacher and pupil talk, plus derived quantities

Teacher	Lesson duration	# of questions per lesson by		Absolute time per lesson of		Actual lesson duration	Ratio teacher to pupil talk	Share of whole-class dialogue	Share of high-lev. questions
		teacher	*pupils*	*teacher talk*	*pupils talk*				
H.	45	44	3	17:56	05:13	38:53	3.4	59%	46%
	90	52	18	19:26	21:30	84:19	0.9	49%	54%
W.	45	44	6	21:22	06:46	37:15	3.2	75%	41%
	90	83	14	35:59	11:46	76:34	3.1	63%	46%
P.	45	67	11	26:57	04:04	42:41	6.6	72%	48%
	60	94	12	36:31	05:36	57:03	6.5	74%	54%
S.	45	14	3	11:52	05:28	42:26	2.2	41%	57%
	60	39	6	21:50	08:24	57:24	2.6	53%	54%

Note. For clarity, the table displays averages of all three videos of the same teacher and the same lesson duration. The full data set can be found in table 3a in the appendix. Since there are only three values per teacher and per condition, standard deviations are not computed (e.g. for the # of questions). Times are given in (mm:ss) and are rounded to full seconds, other numbers to two digits.

Table 4 shows results concerning the dependent variables W1 and W2. It shows the full data set with each lesson individually, plus averages and differences between lesson durations.

According to table 4, wait-times W1 and W2 within one lesson vary between almost zero and four seconds ($\overline{W1}$ ± SD and $\overline{W2}$ ± SD per individual lesson). Furthermore, for all four teachers and for all lesson durations, "double" average wait-times W1 and W2 for a given teacher per lesson duration (Mean $\overline{W1}$ ± SD($\overline{W1}$) and Mean $\overline{W2}$ ± SD($\overline{W2}$)) are between one and three seconds, and therefore slightly longer than those reported in the literature (for non-sensitized teachers). The differences in these average wait times per individual teacher between different lesson durations (Δ Mean $\overline{W1}$ and Δ Mean $\overline{W2}$) are only fractions of a second. They are for all but three cases smaller than the variances from lesson to lesson of same duration (SD of Mean $\overline{W1}$ and SD of Mean $\overline{W2}$ per lesson duration). Only in the case of teacher S (W1 and W2) and teacher W (only W2) are the duration induced differences comparable (0.5s) or even larger (0.4s and 0.7s) than the variances from lesson to lesson (range from 0.1s to 0.9s). The changes in wait-time for teacher S, one of them positive, one negative, are (marginally) statistically significant (10%-level, Mann-Whithey-U-tests) with large effect sizes (Cohen's d). For teacher W in the case of W2 the effect is not (marginally) significant (because there is an overlap in the ranks, see $\overline{W2}$ for the individual lessons, which is used for the U-test), although the mean difference between 45 and 90

minute-lessons is of the same magnitude as the topic-induced variance and Cohen's d is large. Further analysis shows that the single longest (maximum) wait-times W1 and W2 recorded per teacher range from 13 to 18 seconds, and are therefore of the order of the optimal wait-times, see table 1.

Tab. 4. Dependent variables W1 and W2 of the full data set (n=24 videos). Lessons are displayed individually.

Teacher	Individual lesson	$\overline{W1}$ ± SD	Mean $\overline{W1}$ ± SD($\overline{W1}$)	Δ Mean $\overline{W1}$	$\overline{W2}$ ± SD	Mean $\overline{W2}$ ± SD($\overline{W2}$)	Δ Mean $\overline{W2}$
H.	45_1	2.4 ± 2.5	1.9 ± 0.5	–0.2	1.1 ± 0.7	1.2 ± 0.2	0.1
	45_2	2.3 ± 1.8			1.5 ± 1.9		
	45_3	1.5 ± 1.2			1.1 ± 1.2		
	90_1	1.9 ± 1.9	1.7 ± 0.2		1.4 ± 0.8	1.3 ± 0.2	
	90_2	2.1 ± 2.2			1.4 ± 1.2		
	90_3	1.7 ± 1.6			1.0 ± 1.0		
W.	45_1	2.3 ± 2.4	2.2 ± 0.3	0.2	2.1 ± 2.3	1.7 ± 0.5	–0.5
	45_2	2.7 ± 3.4			1.8 ± 2.2		
	45_3	2.1 ± 2.8			1.2 ± 0.7		
	90_1	3.5 ± 3.2	2.4 ± 0.9		1.4 ± 0.8	1.2 ± 0.2	
	90_2	1.7 ± 2.1			1.0 ± 0.4		
	90_3	2.7 ± 3.6			1.3 ± 1.8		
P.	45_1	2.5 ± 2.4	2.3 ± 0.4	0.2	1.2 ± 0.7	1.1 ± 0.2	< 0.1
	45_2	2.9 ± 2.3			1.2 ± 1.0		
	45_3	2.1 ± 2.2			0.9 ± 0.7		
	60_1	3.1 ± 3.0	2.5 ± 0.5		1.2 ± 0.8	1.1 ± 0.1	
	60_2	2.6 ± 2.8			1.1 ± 0.7		
	60_3	2.1 ± 2.2			1.0 ± 0.9		
S.	45_1	1.2 ± 1.1	1.3 ± 0.1	0.4* *(d=2,0)*	1.6 ± 1.0	1.8 ± 0.6	–0.7* *(d=–2,0)*
	45_2	1.3 ± 0.6			1.3 ± 0.5		
	45_3	1.3 ± 0.5			2.4 ± 4.5		
	60_1	1.9 ± 1.7	1.7 ± 0.3		1.1 ± 0.7	1.1 ± 0.1	
	60_2	2.2 ± 3.4			1.1 ± 0.4		
	60_3	1.7 ± 2.2			1.0 ± 0.7		

Note. All times are given in seconds and rounded to one decimal.
*p < 0.1 (10%-level)

5. DISCUSSION

We will first discuss the general data setting and then conduct a case-wise, exploratory discussion.

Concerning the general data setting, all videos feature many questions by the teachers, so that an average wait-time can sensefully be computed. In all videos, the whole-class dialogue occupies roughly 40% to three quarters of actual lesson time, which is typical for German physics instruction (see background). So for the aim of this study, suitable and typical videos were chosen.

Concerning the present case here, the share of whole-class dialogue time per lesson varies considerably from pre to post condition for teachers H, W and S. This might be an indicator for a different lesson structure. Also, there are changes with respect to teacher dominance for teachers H and S, although the pupils still ask remarkably few questions compared to the teachers. There is not much change with respect to wait-times W1 and W2. Wait-times are still away from optimal, especially in view of the substantial share of high-level questions (around 50%).

We will now begin a case-wise (teacher-wise) exploratory discussion incorporating findings of the case studies by Wackermann and Hater (2016) and Hausen et al. (2016). The full use of all variables and other background data (such as the amount of questions) sheds some light on the individual teacher's use of the extra lesson time.

Teacher H shows a big increase in pupil questions from 45 to 90 minute lessons. In the shorter lessons, the pupils of teacher H ask on average three questions, in the longer (double) lessons they ask eighteen. In an interview (Hausen et al., 2016), the teacher reveals that she makes use of the extra lesson time by opening lessons with "pupils asking questions". The pupils apparently make use of this possibility. Questions they ask are not necessarily connected with the current lesson series topic, but may stem from everyday encounters (i.e. the teacher is not prepared for these questions). For such circumstances it may seem surprising that *wait-time W2*, which also measures the time before the teacher gives an answer, stays rather short with one second. In the interview, the teacher acknowledges often not being able to fully answer such unexpected questions. Results from a pupil questionnaire (Hausen et al., 2016) show a medium-size significant decrease in ability to follow the instruction. So the pupils do not necessarily see the opportunity to ask questions as helpful for the rest of the normal instruction. Another striking finding for teacher H is that the pupils have a four-fold increase of their talk time thus changing the instruction from *teacher-dominated* (3:1) to an almost balanced time-share of the whole-class dialogue. This may indicate a change in the role or the aims of the teacher. Furthermore, a reduced share of whole-class dialogue per lesson from 59% to 49% from short to long lessons may indicate a change in the *structure* of the lessons. Video analysis (Hausen et al., 2016) reveals that the extra lesson time not invested in pupil talk is mainly used for extended pupil lab work.

The share of high-level questions increases as the lessons get prolongued. This could indicate that the cognitive level is a bit higher for the longer lessons. Video analysis (Hausen et al., 2016) further reveals only one learning goal per lesson regardless of duration. *Wait-time W1*, however, stays the same from short to long lessons. To sum up, teacher H seems to invest the extra (double) lesson time in encouraging pupil questions and giving them a greater time-share within whole-class dialogue (reduced *dominance* of the teacher), and by extending pupil lab work (different *structure*). This is especially highlighted because by chance one short and one long videotaped lesson covered the exact same topic and lesson goal. So it seems as if teacher H under the condition of double-lessons expands individual 45-minute lessons into double-lessons while investing the extra time in a good pupil-teacher relationship. This is in accordance with the interview.

For teacher W's instruction many of the calculated measures double as the lesson duration changes from 45 to 90 minutes: number of teacher questions, number of pupil questions, time of pupil talk, loss of actual lesson duration compared to nominal lesson duration. Some aspects show variations: The *structure* of the lessons seems to be somewhat different because of reduced share of whole-class dialogue per lesson. And there is a large (Cohen's d), yet statistically not significant, decrease in *W2* from 1.7s to only 1.2s. Rather stable remain the *share of high-level questions* (comparatively low), and the *teacher wait-time W1*. Teacher W's *dominance* is also stable for lessons of different duration. In the interview (Hausen et al., 2016), teacher W states that he uses the extra lesson time mainly for increased exercising. This coincides with the decreased time share of whole-class dialogue per lesson. The pupils state that in the longer lessons they can better follow the instruction (Hausen et al., 2016). Teacher W also explains that he was class teacher under both conditions, therefore frequently loosing or addressing actual lesson time for administrative issues (on average eight minutes in the case of 45 min lessons). Video analysis reveals that there is more loss of actual lesson time due to more classroom-disturbances in the longer lessons, and that teacher W basically teaches the content of two 45 minute lessons after another (Hausen et al., 2016). In summary, teacher W under both conditions follows a classical content-developing style with many relatively low-level questions. In "double-lessons", he basically couples two 45 min lessons after another, therefore many values, e.g. the number of teacher questions, just double. However, at least in the "double-lessons" under investigation here he seems to attempt to prevent growing classroom disturbances by tightening the guidance which is visible in (the non-significantly) reduced wait-time W2. Nevertheless, the pupils seem to be able to follow teacher W's instruction better in the longer lessons possibly due to a decrease in whole-class dialogue and an increase in pupil exercises.

Teacher P's instruction is heavily guided by teacher questions and features a heavy dominance of teacher over pupil talk. Wait-times W1 are on average under both conditions around two to three seconds, which is the highest

in this group of teachers, wait-times W2 under both conditions last on average only one second. Wackermann and Hater (2016) report, and the teacher agrees in the interview, that he makes use of the extra lesson time by adding more learning phases, e.g. adding a reflection phase towards the end of a 60 min-lesson. Since the addition of learning phases, which Wackermann and Hater (2016) characterize as increased educational variety, does not alter the share of whole-class dialogue (around three-quarters of lesson time for both conditions), the dialogue structure of lessons remains unchanged. Maybe in accordance with potentially demanding learning phases, the longer lessons feature a higher share of high-level questions. To sum up, a rather long W1 combined with a rather short W2 may indicate that teacher P knows how to run his teaching style. He drives his instruction with questions. In fact, the pupils state that they can follow P's instruction well under both conditions (Wackermann & Hater, 2016; possibly ceiling effect). Teacher P seems to make full use of the extra lesson time by completing the learning processes with the addition of reflection phases that formerly did not fit into his shorter 45-minute lessons.

Teacher S features the fewest questions of all teachers under both conditions. The share of high-level questions is quite high which suggests a more demanding *level of questions* and tasks, especially for the 45 min lessons.The ratio of teacher to pupil talk (*dominance*) is comparatively low for the 45 min lessons and increases somewhat for the longer lessons. His instruction under the condition of 45 min features the smallest amount of whole-class dialogue, although the share is substantially higher for the longer lessons indicating a change in *lesson structure*. All findings so far suggest some deep structure change from 45 to 60 min lessons for teacher S. And indeed, video analysis and an interview (Wackermann & Hater, 2016) reveal that while under the condition of 45 minutes his instruction only features inquiry pupil lab-work, naturally cutting down on whole-class dialogue, under the condition of 60 minutes one of the lessons shows direct instruction/concept building (lesson S 60_3, see table 3a) using whole-class dialogue and a content-developing questioning style as method of pursuit. Concerning *wait-times W1* and *W2* teacher S shows variations going both ways: A significant and large prolongation of W1 and a significant and large reduction for W2. And the wait-times are relatively short. The pupils state that they can follow S's instruction well under both conditions (Wackermann & Hater, 2016; possibly ceiling effect). Since for teacher S under the condition of 45 min lessons (and to a lesser degree also for 60 min lessons) questions and consequently wait-times seem not to play such a role as for other teachers, wait-time is maybe not the primary variable of choice to assess changes in cognitive activation of pupils for different lesson durations. And yet, for teacher S questions and wait-times in view of the new lesson structure/aims now under the condition of longer lessons may play a more important role.

In three of the four cases (teachers H, W and S) lesson prolongation leads to a somewhat different lesson concept, as indicated by altered shares of whole-class dialogue. Teacher P also makes changes by adding new (and potentially demanding) learning phases, but this is not reflected in altered shares of whole-class dialogue. Teacher S adds direct instruction/concept building to inquiry-based learning. Both teachers teaching 60-min lessons make full use of the extra lesson time as indicated by the actual lesson duration. Both teachers teaching 90-min lessons show a substantial loss of actual lesson time. Teacher W tries to teach two 45 min lessons after another which seemingly causes increasing classroom disturbances, whereas teacher H expands single lessons to double-lessons handing over substantial amounts of class-time to the pupils (pupil questions, extended lab work, loss of actual lesson time). All in all it seems as if 90 minutes may be too long to sensefully conduct instruction, to keep up the flow and concentration of the pupils with teacher H intuitively reducing the demands and teacher W struggling with keeping them up. Both teachers teaching 60-min lessons seem to be able to use the extra time sensefully – teacher P by concluding learning processes within one lesson, and teacher S by complementing inquiry-work with direct instruction.

The teachers under investigation show slightly longer wait-times than other teachers reported in literature. This suggests that the four teachers under investigation conduct an instruction that is slightly more activating than other teachers' instruction (Rowe, 1974; Tobin, 1987). However, the average wait-time is still away from optimal, especially in view of the relatively large share of high-level questions (around 50%). And yet, positively speaking, the maximum recorded wait-times, not displayed here, for all teachers are in the range of the optimal wait-times for high-level questions. So the potential is clearly there!

Another point of discussion concerns the not-so-satisfying interrater agreement for the estimation of the question levels (83%). This is largely due to the fact that many of the questions observed here avoided stating explicitly the intended pupil action, see example in table 2, so the Bloom levels have to be guessed. This is in accordance with other research (Krabbe, Bezold, & Fischer, 2016), which shows that by asking questions, many times the action verb for the pupils does not get expressed. As a result, pupils maybe pick their choice and possibly answer at a lower level than intended. This may reconcile the substantial amount of high-level questions, more typical would be a share of only 25% high-level questions (Borich, 2004), with wait-times more suitable/optimal for low-level questions. It may also reconcile the fact why some literature could not find the expected correlation between average question level and student achievement (Hattie, 2009, p. 182).

6. CONCLUSIONS AND LIMITATIONS

With the study presented here, we intended to explore the effect of lesson prolongation on teacher wait-times. Although longer lessons reduce time pressure and therefore should in principle allow for longer wait-times, the

longer lessons under investigation here do show significantly varied wait-times with large effect sizes only for one teacher. In addition, those variations are not substantial enough to reach optimal wait times – and go both ways.

Concerning limitations, the study here made use of a contingency sample of classroom videos. This includes the fact that the lessons compared (short/long) for a given teacher were from different calendar years with different pupils and foremost were not about the same topic (with one exception by chance for teacher H). However, it could be shown in the results section that the topic-(lesson) induced variance for a given teacher was in most cases larger than the duration-induced difference actually under investigation. So, the main result here that there are only very few changes to wait-time due to lesson prolongation seems to be valid. Another limitation concerns the generalizability. Due to the rather small sample size of four teachers from two different schools, the results are not generalizable. However, the aim of this study was only to set up an evidenve-based hypothesis. In light of this study aim, a total of 24 videos provides substantial evidence.

Another limitation concerns the fact that only the question level of the teachers was estimated, not the answer level of the pupils. As pointed out in the discussion, the relatively high share of high-level questions reported here may not coincide with a high answer level of the pupils. In this study, teacher wait-time was investigated because it is related to the cognitive activation of the pupils. However, the latter was not investigated directly here. Also, the estimation of the question level of the teachers is limited by a crude differentiation in only two levels and by differences in the interrater-agreement. So results concerning the question level should be seen with caution.

Also not investigated in this study was the question level by the pupils. As Dillon (1998) points out, observation studies have shown that pupils ask remarkably few questions, which can be confirmed by this study, and even fewer in the search for knowledge. It can only be guessed that only a fraction of those observed here are in the search for knowledge.

As an outlook, future research maybe should shift the focus away from the question level of the teachers to the answer and question levels of the pupils. Possibly even asking questions is to be questioned, and the focus should shift entirely to pupil activation. After all, it's not what the teachers do – it's what the students do in class that matters (Adams & Slater, 2003, p.1).

In conclusion, and answering the research question, teacher wait-time seems more likely to be a stable characteristic of the individual teacher. Wait-time improvement seemingly does not happen automatically with lesson prolongation, but probably needs sensitization and additional training. The findings here remind a bit on the long-lived and well researched findings on the for many people surprisingly small effects of classroom size on achievement (e.g. Hattie, 2006; Hattie 2009, p. 85). Common discussions include that "the effects of reducing class size may be higher on teacher and student work-related conditions.." than on actual achievement, and common explanations include "teachers not optimizing opportunities presented by having fewer students" (Hattie, 2009, p.86). Applied to the situation of prolongued lesson durations, it may be that there are substantial effects on relieve from too much unrest in a school day (also reported by Wackermann & Hater, 2016, and Hausen et al., 2016) e.g. by teachers spending the extra lesson time on improving teacher and student work-related conditions, see for example teacher H. However, reduced time pressure should in principle allow for longer wait-times and thus a shift in teaching style. This can be seen e.g. in parts with teacher S, who also demonstrates further deep changes of instruction, along with teacher P and lesser with teacher W. Seemingly, teachers need extra sensibilisation and training in order to use these advantages constructively. This coincides with Zander et al. (2014), where the longer prolongation (45 to 90 min) only in combination with a training caused changes in instruction that had an impact on student learning. And Wackermann and Hater (2016) report that the expertise for such sensibilisation and training concerning longer lessons may be present well within the individual school.

ACKNOWLEDGEMENT

Special thanks go to the participating teachers and their pupils for enabling this study!

Also, we would like to thank the anonymous reviewers for their thoughtful and constructive comments!

And finally, we would like to thank the editors of PriSE for their helpful support during the review process!

REFERENCES

Ackermann, S. (2011). *Klassengespräch im Mathematikunterricht* [Whole-class dialogue in mathematics lessons]. Kassel, Germany: University press.

Adams, J. & Slater, T. (2003). *Learner-centred astronomy teaching: Strategies for Astro 101*. Upper Saddle River: Pearson Education Inc.

Begehr, A. (2005). *Teilnahme und Teilhabe am Mathematikunterricht* [Participation in math instruction] (Doctoral dissertation). Retrieved from Freie Universität Berlin. http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000001675.

Bloom, B. S. (Ed.) (1974). *Taxonomie von Lernzielen im kognitiven Bereich* [Taxonomy of cognitive learning goals]. Weinheim, Germany: Beltz.

Borich, G.D. (2004). *Effective teaching methods*, (fifth edition). New Jersey: Pearson Merrill Prentice Hall.

Corey, S. (1940). The teachers out-talk the pupils. *School Review, 48*, 745–752. doi:10.1086/440596.

DESI-Konsortium (Ed.). Eckhard Klieme, Wolfgang Eichler, Andreas Helmke, Rainer H. Lehmann, Günter Nold, Hans-Günter Rolff, Konrad Schröder, Günther Thomé, Heiner Willenberg (2006). *Unterricht und Kompetenzerwerb in Deutsch und Englisch – Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International (DESI)* [Instruction and competence growth in German and English – central findings oft he DESI-study]. Frankfurt: Deutsches Institut für Internationale Pädagogische Forschung.

Dillon, J. T. (1988). The remedial status of student questioning, *Journal Of Curriculum Studies, 20* (3), 197–210, DOI: 10.1080/0022027880200301.

Eliasson, N., Karlsson, K. G., & Sörensen, H. (2017). The role of questions in the science classroom – how girls and boys respond to teachers' questions. *International Journal of Science Education, 39*(4), 433–452. doi:10.1080/09500693.2017.1289420.

Hattie, A.C. (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.

Hausen, C., Wackermann, R., & Krabbe, H. (2016). Einfluss von 90-minütigen Unterrichtsstunden auf die Qualität von Physikunterricht [Effect of 90 min lessons on quality of physics instruction]. In: C. Maurer (Ed.), *Authentizität und Lernen – das Fach in der Fachdidaktik. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Berlin 2015* (pp. 602-604). University of Regensburg.

Haynes, H. (1935). *The relation of teacher intelligence, teacher experience and type of school to type of questions* (Unpublished doctoral dissertation). George Peabody College for Teachers, Nashville, TN.

Heinze, A., & Erhard, M. (2006). How much time do students have to think about teacher questions? An investigation of the quick succession of teacher questions and student responses in the German mathematics classroom. *Zeitschrift für Didaktik der Mathematik, 38*(5), 388–398.

Helmke, A. (2009). *Unterrichtsqualität & Lehrerprofessionalität* [Quality of instruction and teacher profession]. Seelze-Velber, Germany: Kallmeyer/Klett/Friedrich.

Ingram, J. & Elliott, V. (2016). A critical analysis of the role of wait time in classroom interactions and the effects on student and teacher interactional behaviours. *Cambridge Journal of Education, 46*(1), 37–53. doi:10.1080/0305764X.2015.1009365.

Krabbe, H., Bezold, T., & Fischer, H. E. (2016). Sprachhandlungen im Physikunterricht [Discourse functions in physics instruction]. In: C. Maurer (Ed.), *Authentizität und Lernen – das Fach in der Fachdidaktik. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Berlin 2015*. (pp. 202–204). University of Regensburg.

Labudde, P. & Duit, R. (2007). Zum Design einer binationalen Videostudie zum Physikunterricht [Design of a bi-national video-study on physics instruction]. In *D. Höttecke (Ed.), Naturwissenschaftlicher Unterricht im internationalen Vergleich (pp. 631–633)*. Münster, Germany: Lit.

Leitz, I. (2015). *Motivation durch Beziehung* [Motivation through relation]. Berlin, Germany: Springer.

Li, W. S S. & Arshad, M. Y. (2014). Wait-Time and Multiple Representation Levels in Chemistry Lessons. *Malaysian Online Journal of Educational Sciences, 2*(2), 45–53, http://moj-es.net/volume/volume-02-issue-02.

Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K. & Klieme, E. (2007). Gleicher Unterricht – gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht [Same instruction – same opportunities for all? Distribution of pupil talk in whole-class instruction]. *Unterrichtswissenschaft, 35*(2), pp. 125–147.

Patrick, A. & Urhievwejire, E. (2012). Is Soliciting Important in Science? An Investigation of Science Teacher - Student Questioning Interactions. *International Education Studies, 5*(1), 191–199. doi:10.5539/ies.v5n1p191.

Rowe, M. B. (1974). Wait-time and rewards as instructional variables, their influence on language,

logic an fate control: Part one – wait-time. *Journal of Research in Science Teaching*, *11*(2), 81–94.

Rössler, P. (2005). *Inhaltsanalyse* [Content analysis]. Konstanz, Germany: UVK/UTB.

Seidel, T., M. Prenzel, R. Rimmele, I. M. Dalehefte, C. Herweg, M. Kobarg & K. Schwindt (2006). Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie. [Views into physics instruction. Results from the IPN video-study]. *Zeitschrift für Pädagogik*, *52*, 799–821.

Seidel, T. & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning & Instruction*, *16*, 228–240.

Stender, A., Geller, C., Neumann, K. & Fischer, H. E. (2013). Der Einfluss der Unterrichtstaktung auf die Strukturiertheit und Abgeschlossenheit von Lernprozessen [The effect of the segmentation of instructional time on the structuredness and completeness of learning processes]. *Zeitschrift für Didaktik der Naturwissenschaften*, *19*, 209–232.

Tobin, K. (1987). The role of wait time, cognitive level learning. *Review of Educational Research*, *57*(1), 69–95.

Wackermann, R. , Trendel, G. & Fischer, H. E. (2010). Evaluation of a Theory of Instructional Sequences for Physics Instruction. *International Journal of Science Education*, *32*(7), 963–985.

Wackermann, R. & Hater, J. (2016). Der Einfluss der Stundenlänge (45 vs. 60 Minuten) auf ausgewählte Aspekte der Unterrichtsqualität im Physikunterricht am Gymnasium. - Eine Prä-Post-Untersuchung mit zwei Lehrkräften. [The effect of lesson duration (45 vs. 60 minutes) on selected aspects of quality of physics instruction in upper level secondary school – a pre-post comparison with two teachers]. *Perspectives in Science*, *10*, 1–12. doi:10.1016/j.pisc.2015.12.009.

Zander, S., Krabbe, H. & Fischer, H.E. (2014). Mehr lernen durch eine bessere Sequenzierung des Physikunterrichts bei verlängerter Unterrichtstaktung [More learning through better sequencing of physics instruction with longer lessons]. *Abstract für die 2. Jahrestagung der Gesellschaft für Empirische Bildungsforschung (GEBF)*, Frankfurt.

APPENDIX

Appendix 1:

Tab. 3a. Control variables of the full data set (n=24 videos). Each lesson is displayed individually. Table 3 in the main article is a condensed version of this table.

Teacher	Individual lesson	# of questions by teacher per lesson	# of questions by pupils per lesson	Absolute time of teacher talk per lesson	Absolute time of pupil talk per lesson	Actual lesson duration	Ratio teacher to pupil talk	Share of whole-class dialogue per lesson	Density of teacher questions[a]	Share of high-level questions
H.	45_1	29	1	17:16	05:15	38:56	3.3	58%	1.3	62%
	45_2	37	2	14:10	04:12	38:10	3.4	48%	2.0	38%
	45_3	66	5	22:21	06:12	39:33	3.6	72%	2.3	44%
	90_1	53	13	15:09	10:59	88:44	1.4	29%	2.0	53%
	90_2	50	20	22:20	19:41	79:41	1.1	53%	1.2	38%
	90_3	54	22	20:49	33:49	84:32	0.6	65%	0.9	54%
W.	45_1	73	9	15:31	04:41	33:18	3.3	61%	3.6	37%
	45_2	23	1	23:07	08:34	34:52	2.7	91%	0.7	39%
	45_3	35	7	25:29	07:03	43:36	3.6	75%	1.1	54%
	90_1	102	16	26:10	10:40	82:33	2.5	45%	2.8	52%
	90_2	76	20	40:02	08:08	66:50	4.9	72%	1.6	46%
	90_3	71	7	41:44	16:31	80:18	2.5	73%	1.2	38%
P.	45_1	54	8	19:57	02:22	41:27	8.4	54%	2.4	44%
	45_2	86	10	26:11	06:27	43:05	4.1	76%	2.6	51%
	45_3	60	14	34:42	03:24	43:31	10.2	88%	1.6	52%
	60_1	97	6	36:11	04:03	60:38	8.9	66%	2.4	45%
	60_2	89	14	38:03	06:11	54:44	6.2	81%	2.0	54%
	60_3	96	15	35:19	06:35	55:47	5.4	75%	2.3	64%
S.	45_1	18	3	12:42	07:59	42:07	1.6	49%	0.9	61%
	45_2	13	4	07:54	06:43	40:06	1.2	36%	0.9	62%
	45_3	10	3	14:59	01:43	45:05	8.7	37%	0.6	40%
	60_1	26	7	19:16	08:11	57:42	2.4	48%	0.9	62%
	60_2	39	3	25:01	04:55	58:18	5.1	51%	1.3	51%
	60_3	51	9	21:14	12:05	56:12	1.8	59%	1.5	53%

Note. Times are given in (mm:ss) and are rounded to full seconds, other numbers to two digits.
[a]Per dialogue time (# per min.)