MASTER'S THESIS

On

An Approach to Measure the Explaining Quality of Online Explanatory Videos

Institute of Science Education
Physics Education Department
University of Bremen
Otto-Hahn-Allee 1
28359 Bremen

Submitted on August 18th, 2015

by

Cord H. Peters

Assessors

Dr. rer. nat. Christoph Kulgemeyer

Prof. Dr. rer. nat. habil. Horst Schecker

Faculty 1 - Physics and Electrical Engineering

University of Bremen, Germany

This master's thesis was submitted by:

Cord Heinrich Peters
Wiesenstraße 10
28844 Weyhe
Germany

Matriculation Number 2145836

cord.peters@uni-bremen.de

049 (0)162 56 14 150

Acknowledgements

I would like to thank my mom, my family, and everyone else I have so far been allowed to call a friend of mine.

I also thank my assessors Dr. Christoph Kulgemeyer for his excellent support, his candid advice, his open mind, his precious time and his expertise as well as Prof. Dr. Horst Schecker for his guidance, his insights, his encouragement and his abundance of patience which, and more, I have been granted over the years.

Extended gratitude must be awarded to Elisabeth Tomczyszyn who worked out and finalized the categories to classify different instruments and methods of explaining in her previous work (which again is based on previous work by Prof. Dr. Horst Schecker and Dr. Christoph Kulgemeyer) and thus provided the basis for this paper.

Genial thanks have to be given to Dr. Anne-Marlene Blechschmidt and Dr. Monika Breitzke for allowing me to conduct polls in their seminars, where time is always short and of the essence.

Many special thanks need be appertained to Antje Mühlenstedt-Meko, foreign language teacher at KGS Leeste, who did not mind extended telephone calls long past midnight.

Last but not least, my appreciation goes to Anja Hübner, Joanna Hoins and Hauke Schmidt whose criticisms and ideas have enriched this paper immensely.

THANK YOU ALL

Universität Bremen

Diese Erklärungen sind in jedes Exemplar der Bachelor- bzw. Masterarbeit mit einzubinden.

Name: **Cord H. Peters** Matrikelnr.: **2145836**

Urheberrechtliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.
Alle Stellen, die ich wörtlich oder sinngemäß aus anderen Werken entnommen habe, habe ich unter Angabe der Quellen als solche kenntlich gemacht.

Datum Unterschrift

Erklärung zur Veröffentlichung von Abschlussarbeiten

Die Abschlussarbeit wird zwei Jahre nach Studienabschluss dem Archiv der Universität Bremen zur dauerhaften Archivierung angeboten.
Archiviert werden:

1) Masterarbeiten mit lokalem oder regionalem Bezug sowie pro Studienfach und Studienjahr 10 % aller Abschlussarbeiten
2) Bachelorarbeiten des jeweils der ersten und letzten Bachelorabschlusses pro Studienfach und Jahr.

(◉) Ich bin damit einverstanden, dass meine Abschlussarbeit im Universitätsarchiv für wissenschaftliche Zwecke von Dritten eingesehen werden darf.

( ) Ich bin damit einverstanden, dass meine Abschlussarbeit nach 30 Jahren (gem. §7 Abs. 2 BremArchivG) im Universitätsarchiv für wissenschaftliche Zwecke von Dritten eingesehen werden darf.

( ) Ich bin nicht damit einverstanden, dass meine Abschlussarbeit im Universitätsarchiv für wissenschaftliche Zwecke von Dritten eingesehen werden darf.

Datum Unterschrift

Preface

This work arose during my studies at University of Bremen, Germany. It is part of my final exam which took place from April to August 2015 for my master's degree. My subjects are physics and mathematics, whereas this work is related to physics education.

Abstract

The objective of this master's thesis is to evaluate the explaining quality of 37 freely accessible online explanatory videos on Kepler's Laws from the video sharing website YouTube by means of an established category catalogue, respectively to validate the application of a measure based on the category catalogue on YouTube videos. The category catalogue is provided by Elisabeth Tomczyszyn. The need to determine the explaining quality of online explanatory videos becomes clear when comparing the number of users of such videos and the available surveys that research explaining quality of explanatory videos: On the one hand, there are a lot of users who use explanatory videos on a regular basis (Wolf & Kratzer, in press), but, on the other hand, there are only few studies that thoroughly investigate explaining quality found in YouTube videos - despite the fact that some journalists even talk about the biggest turnover since Humboldt's university reforms (ZEIT, 2015b). The measure consists of 45 categories, which emerged both inductively and deductively, and it is optimized for ten-minute-long expert-novice-dialogues. The application of the measure on explanatory videos published on YouTube is investigated via concurrent and predictive validity. The former is determined by rank correlations between the rank assigned by quantitative frequency analysis, which is based on awarded category points (CPs) that are allocated once at the occurrence of a subcategory and other surface features such as **Views**, **Likes** and **Comments**, which are categorized by qualitative content analysis according to Mayring. Determining the predictive validity is achieved by comparing learning impacts of a total of 27 students from University of Bremen by means of two genuine YouTube explanatory videos in a pre-post-test design, whereat the video is shown once between the two tests. The measure is both capable of maintaining its high independence of the raters ($\kappa = 0.860$) and its reliability ($\alpha = 0.690$). Moreover, the measure correlates moderately with the number of content-related comments ($\rho = 0.42$, N = 37) and a little stronger yet with relevant comments that are given to

explanatory videos published in a private context. Furthermore, the measures shows that it is basically capable of predicting learning impacts. The findings of this work thereby endorse the possibility to apply the measure, which derived from a dialogue between two active parties, on online explanatory videos to estimate their explaining quality as well. This permits further research into this rich source of information which has so far been overlooked by the teaching community.

Zusammenfassung

Das Ziel dieser Masterarbeit besteht darin, die Erklärqualität von 37 frei zugänglichen online Erklärvideos zu den drei Kepler'schen Gesetzen vom Videoportal YouTube mit Hilfe eines bereits gesicherten Kategoriesystems zur Bestimmung der Erklärqualität zu beurteilen, beziehungsweise die Anwendung eines auf jenes Kategoriesystem basierten Maßes auf YouTube-Videos zu validieren. Die Notwendigkeit zur Bestimmung der Erklärqualität von online Erklärvideos wird deutlich, wenn die Anzahl der Nutzer solcher Videos den vorhandenen Untersuchungen zur Erklärqualität in solchen Videos gegenüber gestellt wird: Einerseits gibt es viele Anwender, die Erklärvideos regelmäßig nutzen (Wolf & Kratzer, in press), aber nur sehr wenige Studien, die sich tiefgründig mit der Erklärqualität von YouTube-Videos befassen - und das, obwohl manche Journalisten gar von der größten Umwälzung seit Humboldts Reform der Universitäten sprechen (ZEIT, 2015b). Das zu untersuchende Maß ist aus insgesamt 45 Kategorien, die sowohl induktiv als auch deduktiv entwickelt wurden, hervorgegangen und für zehn Minuten dauernde Erklärungen im Dialog optimiert. Das Kategoriensystem wurde von Elisabeth Tomczyszyn zur Verfügung gestellt. Die Anwendung dieses Maßes auf bei YouTube veröffentlichten Erklärvideos wird mit Hilfe von Übereinstimmungs- und Vorhersagevalidität geprüft. Ersteres erfolgt über Rang-korrelationen zwischen dem vom Maß durch quantitative Häufigkeitsanalyse vergebenen Rang, der auf den aufsummierten Kategoriepunkten (CPs) welche beim Vorkommen einer Unterkategorie einmalig vergeben werden, basiert und anderen Oberflächenmerkmalen wie **Views**, **Likes** und **Comments**, die über qualitative Inhaltsanalyse nach Mayring kategorisiert wurden. Die Bestimmung der Vorhersagevalidität erfolgt über das Vergleichen von Kenntniszuwächsen von insgesamt 27 Studenten der Universität Bremen mit Hilfe von zwei originalen YouTube Erklärvideos im Pre-Post-Test Design, wobei das Video zwischen den beiden Tests einmalig gezeigt wird. Es hat sich ergeben, dass das Maß sowohl seine hohe Unabhängigkeit vom Rater ($\kappa = 0.860$) als auch seine Reliabilität

($\alpha = 0.690$) beibehält. Des Weiteren korreliert das Maß moderat mit der Anzahl an inhaltsbezogenen Kommentaren ($\rho = 0.42, N = 37$) und sogar noch leicht stärker mit relevanten Kommentaren, die zu Erklärvideos im privaten Kontext erstellt wurden ($\rho = 0.52, n = 16$). Darüber hinaus zeigt das Maß, dass es Lernzuwächse tendenziell vorhersagen kann. Damit haben die Ergebnisse dieser Arbeit die Möglichkeit bekräftigt, jenes Maß zur Ermittlung der Erklärqualität, welches aus Dialogen zwischen zwei aktiven Parteien hervorging, auch auf online Erklärvideos anwenden zu können, was nun weitere Untersuchungen dieser reichhaltigen und von der Lehrendengemeinschaft bisher möglicherweise übersehenen Informationsquelle gestattet.

"[A]ll physical theories, their mathematical expressions apart, ought to lend themselves to so simple a description that even a child could understand them."

Albert Einstein to Louis de Broglie in Paris in the year of 1927
(Clark, R.W. (1972). *Einstein: His Life and Time*. London: A & C Black.)

Contents

1 Introduction

Within the last couple of years, the Internet has been increasingly penetrating our lives and daily routines. This penetration has been both eased and enabled by the appearance of smart phones and tablet computers, making the Internet ubiquitously available as a place of choice to go to in order to be entertained, stay connected with friends but also to look up information. Hardly does a day go by without being online when new knowledge is usually only a keystroke away as TIME noted in its cover story in 2006, see Figure 1. Wikipedia and YouTube are among the most famous and free gateways that provide access to information and knowledge. YouTube, in particular, is a source of all kinds of videos including explanatory videos and tutorials that may be equally interesting to pupils, students, teachers and educators alike. Even though a lot of pupils and students use those videos to learn, revise, or prepare talks and presentations (Wolf, in press), and plenty of children and adults self-produce and publish videos explaining everything from knitting to Higgs-Boson (Wolf & Kratzer, in press; Wolf, in press), there are hardly any research projects that investigate the quality of those explanations, or their correctness, used in those videos from a didactics point of view. The knowledge of lack of research on YouTube videos is evident to such an extent that even DIE ZEIT, a German weekly newspaper, wisecracked in its July 9 edition:

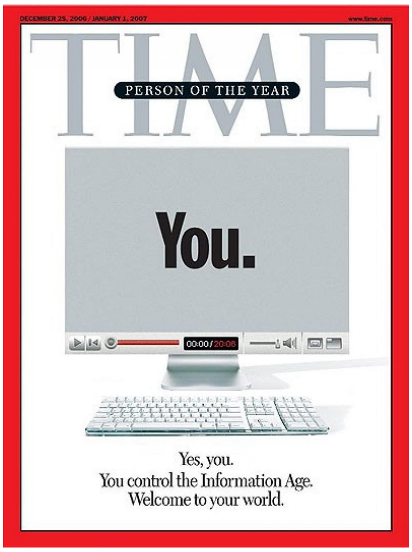


Figure 1: TIME Magazine's cover from December 25^{th}, 2006 foretells YouTube's current success.

> "*Researchers found out that two-year-old toddlers are already able to use a tablet computer. This resulted in an analysis of 200 YouTube videos. To that, one can only say: Great! Science eventually discovered the big treasure of data, which had so far been laying dormant in the Internet embedded in these videos, and has started to hoist it at last*" (ZEIT, 2015a, p. 29, translation by the author).

Since 1995, Merten has already noted that *"as scientists work slowly but thoroughly, it must be feared [...] that the changes of the media landscape take place a lot faster than its research"* (Merten, 1995, p. 18). A closer look into this huge source becomes even more necessary because its use is likely to only expand and become even more commonplace in classrooms, (self-)tutoring lessons and retraining (Wolf, 2015), especially when considering the founder of *The Khan Academy*'s objective, Salman Khan, to enable everyone at any place in this world to get some free education (Khan, 2013). Explanatory videos have become part of plenty of children's, teenagers' and increasingly adults' everyday life but remain mainly unmonitored, resulting in bad performances when used in official settings (Diekmann, Gleiche & Weiner, 2010).

1.1 Topic

This master's thesis investigates the possibility of applying an established measure to gauge explaining skills on explanations of *Kepler's Laws on Planetary Motion* to an audience via the medium of freely accessible YouTube videos. Explaining is usually attributed to teachers and their *professional competence* (Schecker, Parchmann & Krüger, 2014) but its didactics foundation may be expanded to explainers in general such as in a YouTube video producing environment since producers of explanatory videos have to equally anticipate the audience's prior knowledge and adapt their language as teachers have to. Besides, not only is communicating physical contents required by future physics teachers, but it is also a key competence outlined in the formal school curriculum where it means *"to gain and exchange domain-specific information appropriately"* (Senator für Bildung und Wissenschaft, 2006, p. 7, translation by the author) and to gain new knowledge by reading. Communicating in this context also means to work with given presentations, to use appropriate presentations and to choose and use presentations themselves (Senator für Bildung und Wissenschaft, 2006). As a result, communicating domain-specific information is considered a key competence that is to be achieved by pupils attending physics lessons at school. Kulgemeyer and Schecker also point out that *"communication skills are important for scientists as well as citizens, enabling them to participate in the decision-making dialogues of an informed society"* (Kulgemeyer & Schecker, 2013, p. 2236). Consequently, explaining skills used by both professional educators and layman in explanations published on YouTube might overlap concerning certain key aspects that appear essential to give a successful explanation, making it possible to use a measure derived from expert-novice-dialogues be applied on a more diverse and general medium: explanatory videos. Validating this application is part of this work.

The theoretical framework for this validation is derived from physics didactics and psychological theories that deal with communication and explanation techniques. Besides, this work is loosely connected to the large-scale project EWis, respectively ProfiLe-P, which models and investigates how to assess explaining skills (EWis) and whether that skill relates to future physics teachers' factual and educational knowledge (Tomczyszyn, Kulgemeyer & Schecker, 2014).

1.2 Structure

First, the theoretical background is given on topics as explaining and YouTube videos, outlining the latest research projects, followed by a detailed description of methods on how the data are gained and analysed including the materials, processes and manuals used. The findings are presented in Chapter 4, followed by a chapter that discusses and reflects the results as well as it gives an outlook into possible future endeavours.

2 Theoretical Background

In 1986, Brown already experienced that *"[e]xplaining is ubiquitous. It appears in everyday life and in every profession context. Yet, despite its ubiquity there appears to be little research on explaining [...]"* (Brown, 1986, p. 216). That is why this chapter gives information on the explaining skill from different points of view as well as outlines current and latest projects on quantifying the skill sufficiently. It also presents definitions on YouTube's learning videos according to Wolf, and each sub-chapter closes by featuring links to the research questions.

2.1 Explaining Skill

Before defining the explaining skill, the terms **explain or explanation** have to be substantiated first. Bartelborth states that *"explanations lead to an* understanding *of certain occurrences*," or *"explanations are to support our understanding"* (Bartelborth, 2007, pp. 11, 19, translation by the author). The term **explain** usually incorporates the process of giving a verbal explanation and as such includes variables like addressee-adequacy, considering the addressee's prior knowledge, his or her attitudes and skills (Bartelborth, 2007; Brown, 1986). This citation shows that for an explanation to evoke understanding it has to be both viable and compatible for the addressee. That also means that an explanation is basically a precursor to learning or understanding, and in the same manner it underlines the concept that learners themselves have to construct their knowledge by accommodating new facts into their existing knowledge (Duit, 2004). This constructivist learning process can be divided into six cognitive, meta-cognitive as well as motivational elements (Reinmann-Rothmeier & Mandl, 2001):

1. Learning is an active process

2. Learning is a constructive process

3. Learning is an emotional process
4. Learning is a self-controlled process
5. Learning is a social process
6. Leaning is a situational process

Hence, to be considered an effective explanation, it should excite some of the elements above. Bartelborth also states that whether an explanation is defined as rich and varied "*is a subjective aspect. [The explainer] assumes [the addressee's] prior knowledge and his or hers cognitive performance. [The explainer] has used descriptive examples that appeared very neatly to [the addressee]*" (Bartelborth, 2007, p. 11, translation by the author). This addressee-adequacy may also mean that the explainer uses graphic presentations and is able to swap levels of his or her speech, e.g. to switch from a scientific peppered language to everyday language (Kulgemeyer & Schecker, 2013), and to use words which the addressee knows and can relate to (Brown, 1986). These theories are backed up by Wellenreuther's and Brown's work.

Wellenreuther examined various studies on teachers' verbal explanations and states different possibilities to introduce a new field such as activating prior knowledge by asking questions, introducing the new topic verbally and clarify by using examples to discuss students' misconceptions (Wellenreuther, 2013). He also points out that *"language is the most important medium in transferring information"* (Wellenreuther, 2013, p. 178). Moreover, he advises that explanations must consider students' prior knowledge and be part of their living environment. He isolates the following elements that mark a rich and varied explanation (adapted from Kulgemeyer & Schecker, 2013, p. 2241):

- Point out the relevance of the described content
- Repeat the most important points several times
- Use different forms of media
- Start the explanation with regard to the students' prior knowledge
- Have eye contact and be attentive

- Give encouragement and feedback

An obstacle in introducing new topics is identified as the limitations of the student's **working memory** (Wellenreuther, 2013), which can become a more serious problem in complex explanations. Some students may be lost during the explanation process because they do not recall necessary information that the teacher assumes as known. Evertson, Emmer and Brophy found out that *"effective teachers spend more time on explaining, demonstrating, modelling and discussing than making the students perform the tasks in silent work"* (Wellenreuther, 2013, p. 179). Furthermore, students' questions can give valuable hints while they are integrating the new information into their knowledge world. Another tool for helping students integrate new information is identified by Wellenreuther as scaffolding. With offering each student just enough assistance and ideas, based on his or her individual knowledge, scaffolding may improve a student's knowledge structures and attitudes.

Brown investigated the efficiency of teachers' explanations at school based on their research in general as he sees *"[e]xplaining as the core skill in communicating"* (Brown, 1986, p. 201). As a result, his research focusses on addressee-adequacy, or better phrased: on how and when to explain. Brown (1986) describes it as seeing the problem through the eyes of the addressee rather than addressee-adequacy. Therefore, he defines explaining as a process that encompasses the explainer's intention, the process itself and the changes in the addressee's attitude (Brown, 1986). His review of findings of different studies that make up a successful explanation are (adapted from Kulgemeyer & Schecker, 2013, p. 2241):

- Planning: Analyse the topic and adapt a plan according to learner characteristics
- Clarity and fluency: avoid vagueness, use explicit language
- Emphasis and interest: vary gesture, media and material, show emphasis in the use of voice, pause, summaries and repetition
- Using examples: use appropriate examples, positive and negative if applicable
- Organisation: use logical links of words, phrases and sequence patterns
- Feedback: give opportunities to ask questions, ask questions in order to test understanding

One of the main points is feedback as without feedback students are left alone with their own intentions and ideas as well as clarity and interest. (Brown, 1986). Furthermore, students like clear and relevant explanations.

Kulgemeyer and Schecker (2013) point out that even though Wellenreuther and Brown investigated different areas of explaining skills, some elements clearly overlap. Considering all this, however, it is not surprising that explaining is usually seen as a more demanding task compared to listening (Renkl, 1996). Beyond considering the addressee's prior knowledge or misconceptions and addressee-adequacy, a rich and varied science teaching explanation might consist of analogies (Bartelborth, 2007), use of multiple presentations such as texts, pictures, experiments, underlining differences to similar concepts, applying examples and connecting the addressee's experiences; whereas repeating an explanation several times in similar wording, ignoring questions, or describing something as easy and obvious are considered parts of a bad explanation (Schecker, 2007). Moreover, Brown (1986) cited one study *"that better explanatory lessons had more keys and more types of keys. In other words, [the teachers] varied the cognitive demands on the pupils and they used higher levels of cognitive demands"* (p.208). It also shows that explaining something to someone that he or she can fully understand is entangled with how the information is delivered and subsequently allowing the addressee to work with the explanation according to his or her inner construct of reality (Merten, 1995). This perspective of explaining necessitates communication models that see both the explainer and the explainee in an active role. Two of these models are highlighted due to their importance to the measure for gauging explaining skills:

- Merten's communication model published in 1995 sees the communication process as dynamic rather than static, which implies that the same stimuli sent by a medium always lead to the same responses in recipients (Merten, 1995). One basic idea in this model is that *"the reality in which human being live, is a reality which the human being has created himself"* (Merten, 1995, p. 7). Hence, each person creates his or her reality subjectively. One main tool for this constructing is **selectivity**. Merten states that selectivity functions like a "gate-keeper" to prevent an overload caused by to much information (Merten, 1995). This selection process is controlled by interpretation, attitudes, expectations and memories. Representing this point of view, however, makes it possible for the addressee to misinterpret information (Merten, 1995). With adapting the effect of communication as change of knowledge, attitudes or behaviour, Merten uses a common definition of explaining in communication theories, before breaking off by listing disadvantages of the old communication model, such as (Merten, 1995, p. 14, translation by the author):

1. Causality: Cause of effect is stimuli only
2. Proportionality: Effect is dependent on the frequency and strength of stimuli
3. Transitivity: The stimuli is understood exactly as intended by the sender

As a result, he adds along with (1) internal (knowledge, attitudes) and (2) external (values, situation) contexts also (3) expectations and (4) feedback, which influences the addressee's selection by adding favourites for example, to the previous model. Merten (1995) underlines the fact that selectivity itself also changes over time as the addressee's knowledge and attitudes alter. This is part of the dynamic effect of his model.

• Rusch's communication model from 1999 is a constructivist communication model in which communication is understood as an orientation-and interlinking action (Rusch, 1999). In his model, communication may also be unsuccessful, which is a new aspect. His main concern is stressing the fact that communiaction can also be ignored and it is part of the communication act to put attention on a common communicating basis.

To put it in a nutshell, communication models that see the listener, or addressee, in an active role are required to incorporate the listener's choice to either work with the explanation in a constructivist way, or to ignore it due to selectivity or an overloaded working memory. Consequently, the addressee's reactions or feedback help the explainer modify his or her approach, which may set a whole feedback loop into motion.

2.1.1 Definition of Explanations

Explanations can be differentiated into **scientific explanations** and **science teaching explanations** (Bartelborth, 2007); this work focusses on the latter since it is the kind of explanation used in educational environments. The purpose of science teaching explanations is to make topics understandable to a certain audience, and a dominating factor in this process is the consideration of what the audience exactly needs to know in order to grasp a new concept (Kulgemeyer & Tomczyszyn, in press). Brown further distinguishes between three teaching explanations (Brown, 1986, p. 205):

- Reason-Giving: Answering Why-questions. (Why does the volume of a gas decrease as pressure increases?)
- Interpretive: Answering What-Questions. (What are the effects of a high inflation rate on currency?)
- Descriptive: Answering How-Questions. (How can a perpendicular be constructed using compass and ruler?)

On the other hand, scientific explanations can be characterized by Hempel's deductive-nomological model which "*states that an event E that is to be explained (explanandum event) was to be expected due to certain given pre-conditions and laws of nature. E was also highly anticipated, since it was derived deductively from laws G1,...,Gr and the pre-conditions A1,...,An (combined the explanans)*" (Bartelborth, 2007, p. 23, translation by the author). This model has been proven to reconstruct scientific explanations (Bartelborth, 2007). Similarly, "*explanations differ from argumentations insofar as an argument supports a claim, while explanations mainly aim at understanding why and how a phenomenon occurs*" (Kulgemeyer & Tomczyszyn, in press, p. 6, translation by the author). Concluding, the phenomenon itself is indisputable in an explanation, whereas a claim in an argument is not (Bartelborth, 2007).

2.1.2 Measuring Explaining Skills

How to measure something as complex and diverse as the **explaining skill** or **explaining competence**? Tomczyszyn and Kulgemeyer have therefore advanced an assessment method that not only observes the product but also the process of explaining orally and results in 45 categories that describe different levels and aspects of a rich and varied explanation that fulfils the elements mentioned in 2.1. Their summation, if used, is called **measure**, see 3.4.1. Hence, their definition of an explanation slightly differs from the definitions given above but it encompasses the gist of science teaching explanation.

"*The process of explaining consists of addressee-adequate and appropriate compilations and modification of explanations used in class*" (Kulgemeyer & Tomczyszyn, in press, p. 10, translation by the author).

Their assessment method, founded in previous studies on science communication competence (SCC) by Kulgemeyer and Schecker (2013), is based on a controlled test setting in a total of 64 ten-minute expert-novice dialogue situations that are video-recorded for in-depth analysis of the experts' reactions and modifications to prompts delivered by the novice. Their following qualitative analysis provided further inductive categories that supplemented the deductive categories derived from general communication theory, Rusch's constructivist communication model, which sees communication not as a technique of transmitting meanings, but a tool for participant's orientation (Rusch, 1999) as well as Merten's model that puts the addressee in an active role insofar as he or she can interpret perceptions differently according to his or her attitudes, expectations and memories (Merten, 1995), and Kulgemeyer and Schecker's communication model (Kulgemeyer & Tomczyszyn, in press). Those 45 categories derived from their study serve as the basis for measuring explaining skills quantitatively in this work. Refer to 3.3.1 to learn how the categories were attained at length. With Kulgemeyer and Schecker's SCC model being the basis for some of the categories, their model is presented in further detail:

Kulgemeyer and Schecker's model consists of three main parts: the **communicator**, the **addressee** and the **factual content** that is to be communicated (see Figure 2). The communicator, however, can only guess the addressee's prior knowledge leaving him or her two options during the process. First, *"try to meet the assumed requirements of the addressee"* or second, *"strictly follow the structure of the factual content"* (Kulgemeyer & Schecker, 2013, p. 2238). According to constructivism, the addressee assumes an active role insofar as he or she has to decide whether to use the communicated information or not (Kulgemeyer & Schecker, 2013).

Given this model, the communicator can dispose of four **aspects** he or she can change in order to improve his or her offer: (1) **Factual content** describes the chosen topic, (2) **context** specifies the *"factual connection in which the information is presented"* (Kulgemeyer & Schecker, 2013, p. 2239), (3) **code** discriminates between levels of language

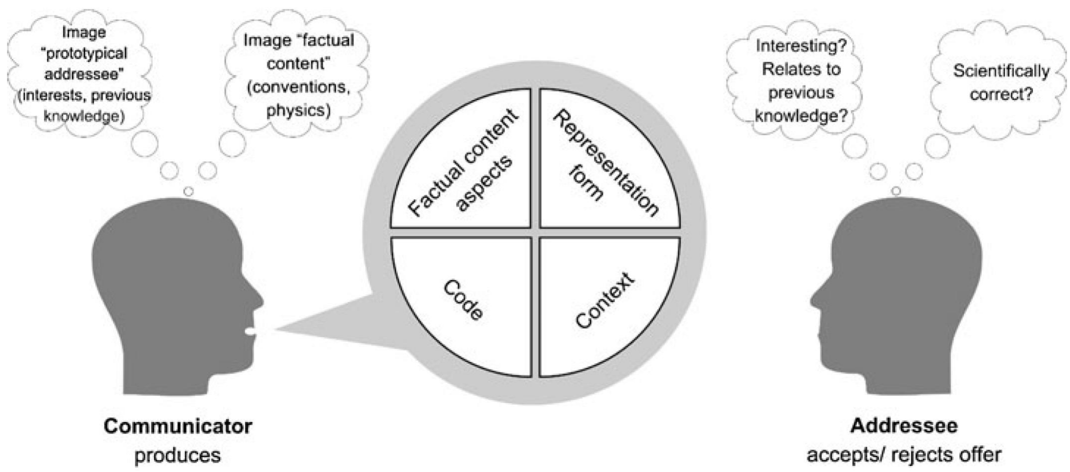


Figure 2: Kulgemeyer and Schecker's (2013) constructivist communication model for science communication.

and (4) **presentation form** characterizes the different means of presentation (Kulgemeyer & Schecker, 2013). This models therefore incorporates both requirements for successfully explaining scientific content in a science teaching environment - addressee-adequacy and subject-adequacy (Kulgemeyer & Schecker, 2013).

The addressee-adequacy as well as constructivism are the main foci of their categories and correspond with Bartelborth's, Brown's, Wellenreuther's and Rusch's as participation in the communication process as the addressee requires attention by observing and listening (Rusch, 1999), active interpretation as well as not to overload an addressee's working memory (Wellenreuther, 2013). It is, above all, the learner's task to use the explainer's offer in from of an explanation to construct his or her view of reality as defined by constructivism (Kulgemeyer & Tomczyszyn, in press). Consequently, *"there is no such thing as a good explanation. What's 'good' for one group may not be good for another"* (Brown, 1986, p. 203). Von Thun (1998) also emphasises the important role the listener takes on during communicating as he or she has to interpret the talkers message on four different levels: relationship, plea, factual content and self-disclosure.
Tomczyszyn and Kulgemeyer's 45-category list, although already tested on recorded expert-novice dialogue environments, has not been tested on published explanatory videos from the video-sharing website YouTube. Whether this is another field of application for their findings is to be investigated in this work.

2.2 Explanatory Videos

YouTube has been an abundant and free source of home-made and self-produced videos as well as rips of blockbuster films and TV series for some 15 years (Wolf, in press). *"For every minute that passes, we upload 300 hours of new content to YouTube"* (TIME, 2015, p. 36). That is why it has been chosen as the primary source for explanatory videos in this research, apart from media centres of broadcasting corporations. To show YouTube's impact on its role in society, Wolf mentions that *"video portals have established themselves, apart from the pure function as entertainment centres, as a kind of [audio-]visual encyclopaedias of both everyday and highly specialized knowledge"* (Wolf, 2015, p. 2, translation by the author). He also found a way to categorize those user generated **learning videos**, whose definitions are picked up by this work.

First, Wolf (2015) discriminates between professionally produced educational or documentary **films** and user generated **videos**, which were made possible by cheap video recording devices such as smart phones and tablet computers as well as free-to-use video portals like YouTube. Further, *"**explanatory videos** are self-made short films which explain how something is done or how something works, or in which abstract concepts are being explained respectively"* (Wolf, 2015, p. 4, translation by the author). Video tutorials and performance videos are sub-genres of explanatory videos: Video tutorials encourage to mimic whole acts whilst performance videos are used to show off successful actions and do not possess any didactic purpose (Wolf, 2015). However, the separation is not selective, e.g. a performance video can turn into a tutorial if explanations are added. Explanatory videos can furthermore be characterized by thematic diversity, creative diversity, informal communication style and a diverse authorship (Wolf, in press).

2.2.1 Advantages & Disadvantages

The advantage of an explanatory video is that learners are able to watch it at home or wherever they feel comfortable (Chandra & Watters, 2012), and they can watch it as long as it takes them to comprehend the concept. Besides, while watching a YouTube video, students are in their familiar position as addressees or learners but additionally, they can directly control their pace of learning by stopping, slowing down, or simply watching some parts of the video again. The addressee is also more thoroughly engaged as he or she gets stimuli via the visual and aural channels. Kleinhanß (2015) says:

> *To work with videos in an educational environment may be especially interesting because image and sound levels and, if necessary, a superimposed text amend each other reasonably. Thus, a complex concept can be communicated both visually, aurally and in a written manner.* (p. 41, translation by the author)

Apart from engaging students more actively, Wolf points out another advantageous factor that explanation videos offer: variety. *"Only since videos were uploaded to YouTube, the necessary wide range of themes has been realized [to provide] a diverse, alternative, partly overlapping amount of videos"* (Wolf, in press, p. 3, translation by the author). This wide range of available videos on the same topic offers many a different point of view, providing viewers with a choice of videos to find the one that suits their learning habit or prior knowledge the most. Also, explanatory video producers do not have to submit to the audience's requirements in a way TV productions have to, consequently widening the range of available videos on a certain topic and as a result improving the accessibility for different parties (Wolf, in press).

A major disadvantage is that most viewers, children, students or adults alike, cannot determine whether the content of an explanatory video is correct or not unless they are experts on the field broached by the video. Another disadvantage worth mentioning is that contrary to Kulgemeyer and Tomczyszyn's expert-novice dialogue, there is no direct interaction between explainer and addressee, making it impossible for the explainer to immediately analyse and modify his or her explanations according to the addressee's reactions. This is a challenge video makers have to face and address in their production process, which Wolf (2015) outlines in a nine-step process. Not only does the video assist the viewer in his or her learning, but the production includes cognitive and meta-cognitive learning strategies for the producer as well, turning the production of an explanatory video into a beneficial learning situation. A production cycle may look like the following:

1. Choosing a topic
2. Preparing the content including securing own understanding, researching background and typical pre-and misconceptions
3. Drawing a form and idea of explaining compromising of didactic and aesthetic

arrangements of the explanation

4. Preparing and presenting a storyboard
5. Preparing the shooting and organizing props, choosing a location and familiarizing with the technology
6. Taking the shoot
7. Post-production including cutting, adding voice-over and embedding music and titles
8. Uploading the finished product
9. Reflecting upon the content, didactics and aesthetics of the video

Some components can be found in the aforementioned categories of the communication model, too, underscoring its learning benefits for producers.

2.2.2 Previous Findings from Explanatory Videos

Explanatory videos are characterized by informal or colloquial language and focus on factual levels to keep production simple and clear (Wolf, 2015) which does not limit but strengthens the importance of factual content variations as proposed by Kulgemeyer and Schecker's communication model. Wolf's investigation of ten explanatory videos reveals that general as well as domain-specific didactics components are included, such as *"structuring, didactic reduction, authenticity, use of examples, utilizing prior knowledge, including negative knowledge and hints to self-monitoring"* (Wolf, 2015, p. 11, translation by the author).

This already hints to the similarities between the basis of explanatory video and expert-novice dialogues and as a result to the 45-category list's new field of application despite its origin from another setting.

In a related study, Wolf choose 120 out of 600 explanatory videos dealing with sports, environmental protection and school-related topics on maths, computer science, sciences and technology primarily produced by six to 13-year-old children, assuming children mostly apply teaching explanations known to them from their parents, peers, school, TV shows and, to a minor extent, other YouTube videos (Wolf & Kratzer, in press). Those videos were then analysed by means of a slightly modified but previously established standardized category catalogue to retain objectivity. Kulgemeyer and Schecker's domain-specific didactics categories were partly included in this catalogue.

It becomes obvious that teaching explanations, in contrast to explanatory videos featuring sports tricks, for instance, are presented in a typical educational setting, similar to that found at school, supporting the above assumption. Also, those videos feature didactic implementations like reductions, revisions and summaries as well as preparing and presenting a storyboard (Wolf & Kratzer, in press).

Although, an exact teaching explanation analysis remains open, it shows that science teaching explanation videos are similar in their didactic preparation and presentation to the expert-novice dialogue, enabling the established 45 categories to be tried on explanatory and arbitrary YouTube videos dealing with science teaching explanations. Additionally, some steps of the nine-step production process similarly focus on both addressee-adequacy and incorporating the addressee's prior knowledge like the communication models on which the 45 categories are based, further supporting the attempt.

2.3 Blended Learning & Edutainment

To complete the display of explanatory videos, a brief description and definition of blended-learning is given - a learning environment in which explanatory videos may be part of the learning experience. Blended learning is, according to Mandl and Kopp's definition *"a variation of face-to-face phases and virtual phases of a learning environment, which uses the advantages of both of these kinds of teaching and learning for an optimal knowledge acquisition"* (Mandl & Kopp, 2006, p. 3, translation by the author) (See Table 1). Clark (2005) says that blended learning is another method to better modify learning environments to the learner's unique needs, meaning a *"more interactive student-centred style"* (Clark, 2005, p. 20, translation by the author). In this context, media are not used to simply beautify or illustrate, but are specifically included to support and promote learning.

Table 1: Disadvantages and advantages of face-to-face and e-learning environments (Garrison & Kanuka, 2004; Reinmann-Rothmeier, 2003).

Face-to-Face:	
Advantage	Disadvantage
• social cooperation	• bound to a place
• direct interaction between educator and learner	• bound to a time
• better communication	• diverse prior knowledge
• discussions arise	• no individual learning pace
E-Learning:	
• independent from place	• no social bonding
• independent from time	• misunderstanding may remain undetected
• determining own pace	• self-organizing required
• methods and didactics are flexible	• educators may loose overview
• responds to different types of learners	
• misunderstanding can occur	

It appears obvious that some face-to-face learning environment disadvantages can be compensated by e-learning components and vice versa. However, new obstacles may arise when introducing an effective blended learning environment like developing a coherent curriculum, choosing appropriate media and addressing a lot of learning channels are only some of the points that must be evaluated by the educators. Also, the dependence on technical and digital devices must be noted and considered. Garrison & Kanuka (2004) underline that those devices must be reliable, easy to use and cutting-edge. Also, the students have to possess knowledge on how to use computers and software to completely use the digital content efficiently and to their advantage (Alexander, 2010). Last but not least, an educator, who took part in Chandra and Watters's survey, pointed out that *"chat-sessions were good and effective. However, they were very labour extensive. Teachers have to find additional time and resources to make this activity happen"* (Chandra & Watters, 2012, p. 638). Additionally, blended learning is not to be confused with pure e-learning elements such as computer-based-training or web-based-training which do not incorporate face-to-face phases. To integrate virtual learning phases, they must be understood under the condition of a constructivist approach, since the learner uses virtual learning phases to acquire knowledge or skills according to his or her own learning habits. The term **blended learning** itself offers a lot of interpretations that is why Mandl and Kopp (2006) underline the following:

> *[I]t is necessary for the learner to be enabled to choose from different approaches to the content. This can be achieved for instance by reading a book, watching a video or participating in an excursion. It is paramount to place the learner at the centre of the considerations.* (p. 8, translation by the author)

These conditions foster the use of science teaching explanation videos due to their same intentions grounded in communication models and how learning takes place. However, Mandl and Kopp could not support blended learning's expectations to reduce time as well as costs and simultaneously increase efficiency, although students accepted blended learning arrangements in general (Mandl & Kopp, 2006). They also point out that the research into blended learning environments is still at the beginning. Moreover, Clark realized that the *"overall impact on student learning was at least neutral"* (Clark, 2005, p. 24).

To fulfil those high expectations set for education in the twenty-first century, the media become crucial, and a high educational demand is imposed on, e.g. the videos. Educators have to decide between low and high quality explanatory videos that are to be included into their learning environment. As educators have to evaluate, if they decide not to produce digital material themselves, which existing videos they deem suitable for their purpose, they decide on the explaining quality of the video. This example shows that standardized evaluating tools that judge the quality of explanatory videos might be considered a help in that decision-making process.

Chandra and Watters, on the other hand, investigated the effect of blended-learning on students' attitudes in a design-based research on 48 students versus a control group of 32 students of the same high school in Australia through quantitative pre- and post intervention and qualitative data from surveys. The web-based learning environment *Getsmart* was theory-based edited and augmented by findings during its application. He found out that participants *"had a positive perception of a blended web-based learning environment, and that such an approach had a positive influence on students' attitude towards their subject"* (Chandra & Watters, 2012, p. 632). The students also improved their knowledge of concepts in physics (Chandra & Watters, 2012).

For the sake of completeness, edutainment is not to be confused with blended learning environments since it is entertainment (such as through games, films or shows) that is designed to teach something (Merriam-Webster, 2015). Thus edutainment focuses on entertainment; examples may include *Sesame Street*, *An Inconvenient Truth*, radio shows, museums and so forth. Edutainment pursues a different educational aspect than explanatory videos, therefore edutainment is merely mentioned in this context to separate it from the former.

2.4 Kepler's Laws on Planetary Motion

In order to assess the impact of explanations from explanatory videos on students' knowledge, the information ought to be entirely or nearly entirely new to the students and not memorized from previous lessons, it was decided on a topic that is not part of the federal school curriculum (Senator für Bildung und Wissenschaft, 2006), but still easy enough to be completely explainable in explanatory videos in a reasonable period of time. Therefore, **Kepler's Laws on Planetary Motion** came to mind since they require no mathematical derivation as Johannes Kepler (1571-1630), a German astronomer, mathematician and founder of geometric optics, entirely founded his laws on observations conducted by Tycho Brahe (1546-1601). Hence, the explanation of his laws focusses on the laws themselves and their possible implications. They are (Bergmann, 2001):

1. Kepler's First Law: The orbit of a planet is an ellipse where one focus of the ellipse is the sun.

2. Kepler's Second Law: A line from the planet to the sun sweeps out equal areas in equal amounts of time.

3. Kepler's Third Law: The period of a planet's orbit squared is proportional to its average distance from the sun cubed.

Additionally, this author's experience shows that astronomy motivates most students, eliminating a possible future obstacle when students participate in the field study to evaluate the learning impact of videos.

The noteworthy implication of the second law is that a planet's speed is not a constant value. Its speed is the slowest at aphelion and fastest at perihelion, the point closest to the sun. The third law is only valid if the objects, e.g. planets or satellites, orbit around the same focus. Therefore, it holds true, for example, either for all the planets orbiting the Sun, or for all the satellites, including the Moon, orbiting Earth. Furthermore, the third law is only an approximation due to the planets' masses, which are generally neglected (Bergmann, 2001).

Known pre- and misconceptions on Kepler's Laws include the perception that planets move around a star in circles at a constant speed, and that all planets have the same orbital period. The Conceptual Physics Workshop conducted by the University of Dallas, Texas, hints at three additional misconceptions, found in its brochure (University of Dallas, 2013, p. 5):

- An object must be at both foci
- All orbits are exactly in the same plane
- Revolution is the same as rotation

3 Methods

This chapter outlines the research question linked to open questions from the previous chapter and the methods used to answer this question. It also illustrates the step-by-step processes and reasons for mathematical techniques applied as well as their limitations. Furthermore, the test instruments, their development, execution and quality criteria are also described.

3.1 Research Question

The overall research question is whether **an established measure for dialogic explaining quality can be used to determine the quality of explanatory videos?**. Both the similarity and analogy between explanations given in explanatory videos and in expert-novice dialogues as stated in (2.2.2), support the use of the proven and as such established 45 categories to gauge explaining quality in explanatory videos with science teaching explanations involved. Likewise, the example given in (2.3) also triggers the research question aiming at supporting educators, teachers and lecturers in their decision-making process to choose suitable explanatory videos that satisfactorily utilize addressee-adequate, prior knowledge-based and rich and varied science teaching explanations. Also, this measure, the used categories summation, could have some correlation to other quality-related criteria provided by the video portal YouTube. The measure then must prove to give similar results in practice, thus justifying the last research question.

Hence, in order to investigate the explanations found on explanatory videos, this three-step process is used, guided by the following overall (question (1)) and the two broken-down research questions (2) and (3):

- (1) **Can an established measure for dialogic explaining quality be used to determine the quality of explanatory videos?**
- (2) **Are there significant correlations between the measure and quality-related criteria such as *Views, Likes, Average View Duration* and *Comments*?**
- (3) **Does the measure predict the learning impact of explanatory videos?**

The term **learning impact** in this context is defined as the *"effectivity of an event which can be described by result testing"* (Schüßler, 2004, p. 43, translation by the author). Impacts are understood as both changes in behaviour or attitude and learning success, which can be tested by achieving the comprehension of, for instance in this case, Kepler's Laws. For this approach, the objective must be known first.

3.2 Design

To assess something as volatile as explaining skill, a category system was compiled from previous researches (see 2.1) and these 45 categories, respectively the points awarded by that category system (see 3.4.1), also called the measure, already prove to be reliable and valid in its expert-novice-dialogue environment(2.1). However, to apply these categories to a more arbitrarily selected set of explanatory YouTube videos of various run-times represents a new application and thus needs verification. To accomplish this, it is decided to first examine **concurrent validity** to additional surface criteria provided by the video distributor YouTube such as **Views, Likes** and **Comments**, which represent the main channels of communication between explainer and addressee, by comparing the rank of a video via Spearman's rank coefficient to the rank awarded by the measure. Second, the measure ought to distinguish between videos with plenty of different categories used versus those with less categories used and probably between video class, say, for videos for classroom or private purposes. The former are thus considered superior whilst the latter inferior. If this obtained rank can be reproduced by students' gain of knowledge from watching those videos assessed by a paper-and-pencil test, it would give hints to the **predictive validity** of the measure. However, predictive validity can only be evaluated if *"the evaluated hypothesis is true and that the criteria* *for validity are measured effectively"* (Wellenreuther, 2000, translation by the author).

In order to finding instruments that are able to answer the aforementioned research questions, they need be properly placed within the realm of subject-didactic researches: The above objective belongs to **basic research** since this work focusses on analysing and reflecting on methods and processes to qualify the explaining skill and learning impact of videos. Hence, the **topic area** includes **learning** which incorporates the consideration of the audience's prior knowledge and addressee-adequacy as the main criteria (Schecker et al., 2014). Placing the research question in the above fields shows that traditional science data analysis does not suffice to generate the required data. For this reason, data collecting methods from social studies are applied such as qualitative content analysis as well as determining reliability and validity measures.

3.3 Data Collection

While some data are provided by the video-sharing portal YouTube, a subsidiary of Google Inc., others are derived from field studies via a paper-and-pencil knowledge test in a pre-post-test design.

Although other video-sharing portals dedicated to teaching or tutoring such as *www.draufhaber.de*, *www.sofatutor.de* or *www.khanacademy.com* were considered but later omitted since they require an account and/or were not available at no charge for the user. Broadcaster's media centres are, for two reasons, another omitted medium: First, most videos are not explanatory videos but films according to their run-time such as documentaries; and second, they do not offer any interactive features other than views or comments and those were limited. Therefore, YouTube is the sole choice, not only due to its huge amount of video material, but also because of its popularity and free access to its video contents as well as its policy to let viewers participate and contribute actively (Welbourne & Grant, 2015).

3.3.1 Data for Concurrent Validity

The data collection for research question (2) took place in a laboratory study via an in-depth analysis of explaining skills from April to May 2015. All other data related to YouTube surface features are from May 2^{nd}, 2015. The videos were found by using the following queries in English and German on YouTube's own search engine as well as YouTube's **Up-Next** feature, resulting in a sighting of more than 100 unique videos:

1. kepler('s) law(s) (explained) / kepler('sche) gesetz(e) (erklärt)
2. kepler('s) law(s) (on) planetary motion (explained) / kepler('sche) gesetz(e) für planetenbewegung (erklärt)
3. planetary motion (explained) / planetenbewegung (erklärt)

A code is then assigned to each short-listed video: The first letter denotes the language (**D** = German; **E** = English), the second letter states the purpose (**P** = private purpose, free time activity; **U** = teaching purpose at universities, schools, colleges; **T** = TV purpose, popular science magazines), the two numbers denote the run-time of the video **rounded to the next full minute**. (If the time is less than ten minutes, a zero is prefixed.) Furthermore, if a video is part of a series, or a code has already been assigned, a small letter starting at the beginning of the alphabet and subsequently moves one letter further on per video, will be added. That letter can be repeated as necessary to register the number of the episode in the series. For example, the video **DP11** is in German, has private background and is less than eleven minutes long, whilst **EU07aaa** is in English, has teaching background, is less than seven minutes in run-time and is the third episode of a series.

With testing of the measure being the objective, knowledge of the population is not necessary and in this case unknown, therefore a deliberate sampling method is justified, applying the search parameters stated above. The only known parameter of the population is that the video explains Kepler's three laws on planetary motion in either English or German.

Statistical Data from YouTube

The data needed for determining concurrent validity are provided by YouTube and are considered surface features such as **Views**, **Likes**, **Average View Duration** and

Comments. Not only are these surface features the only channels to allow communication between the explainer and addressee, but that kind of communication is asynchronous, meaning the communication is timely and spatially separated, resulting in a delay in responses and feedback.

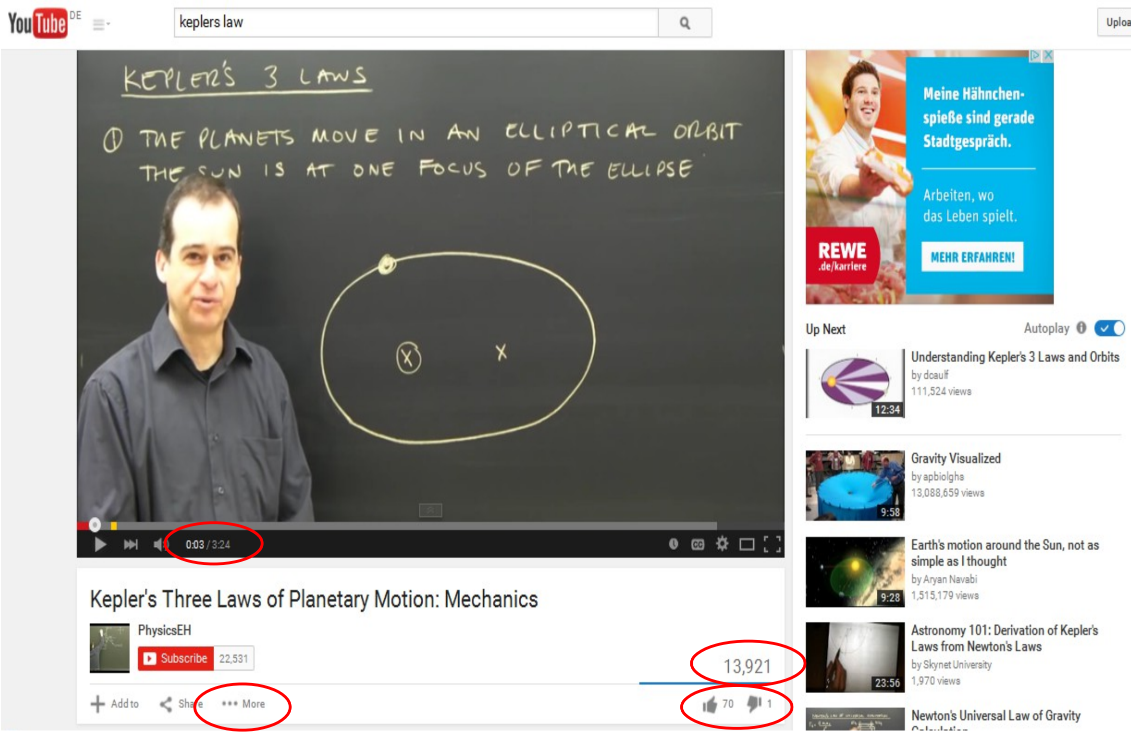


Figure 3: *Views*, *Likes/Dislikes*, run-time and further statistics found on YouTube's user interface. Here the users' original interface of video **EP04p** as of 08/June/2015 is shown as an example, including YouTube's *Up-Next* feature.

- **Views**

The **Views** are among the most easily recognized surface features. They are provided in the bottom right corner of a video in a slightly bigger font (see Figure 3). YouTube's definition of a **Views** is *"a viewer-initiated, intentional play of a video"* (Parsons, 2014) either on YouTube.com or embedded in a third party's website. A YouTube user found out that YouTube counts a **View** if a video is started and watched for as few as five seconds (LaPointe, 2010). **Views** of less than 300 are not audited by the video-sharing provider and as such might be artificially increased. However, whenever the amount approaches approximately 300, YouTube starts verifying the number of unique views and amends the amount as necessary by deleting views coming from the same location.

Until then, the number of views can be manipulated by loading the video over and over again (Parsons, 2014). Furthermore, if a video is of the auto-playing kind due to players used, it will not be considered a "viewer-initiated" action and as thus not cumulated as a **View** (Van, 2013). Also, Google Developers state that cookies are used to accumulate **Views**, allowing counts from different devices or web browsers for the same person (Google Developers, 2015).

With **Views** being the main surface feature and basis for revenue, users are still trying to artificially inflate their amount of views, a service sold by a lot of websites (Welbourne & Grant, 2015). However, YouTube works hard to prevent that by smart algorithms (Parsons, 2014). That means that videos with less than 300 **Views** must be considered more cautiously, but the majority of the authorship of explanatory videos might follow an honest path since they appear to be candidly interested in the viewers' comprehension. Nevertheless, its possibility cannot be fully excluded, either. Moreover, YouTube's parent company, Google Inc., primary objective is to run a profit, thus it tries to promote popular videos even more, turning the **Views** count into, as Welbourne and Grant (2015) call it, a *rich-gets-richer* (p.3) commercialisation.

- **Likes / Dislikes**

Whether a viewer likes or dislikes a video, he or she can voluntarily vote by clicking either a thumb-up or a thumb-down symbol. For this action, an account is necessary to prevent fraudulent manipulations and allowing just one vote per viewer. The cumulative voting can be seen right below the **View** display in the bottom right corner of the video interface. If a user has already voted, the symbols changes to green (thumb-up) or red (thumb-down) to notify the user that he or she has already voted on that particular video.

- **Average View Duration**

This surface feature can be found under **Video Statistics** (see Figure 3) and the button labelled **More**, followed by the tab labelled **Time Watched** (see Figure 4). To gain access to that information no account is necessary, however, the publisher can decide via his or her video settings to make the display of **Average View Duration** public, hence some videos do not show this surface feature. Google Developers say that it is the average length of video playbacks in seconds (Google Developers, 2015).

To compare the **Average View Duration** to the complete run-time of the video, run-time can be found in the left bottom corner of the video interface (See Figure 3). Among YouTube publishers the **Average View Duration** is among the most contested and discussed surface features in blogs as it shows on average how long viewers remain engaged with the video content.

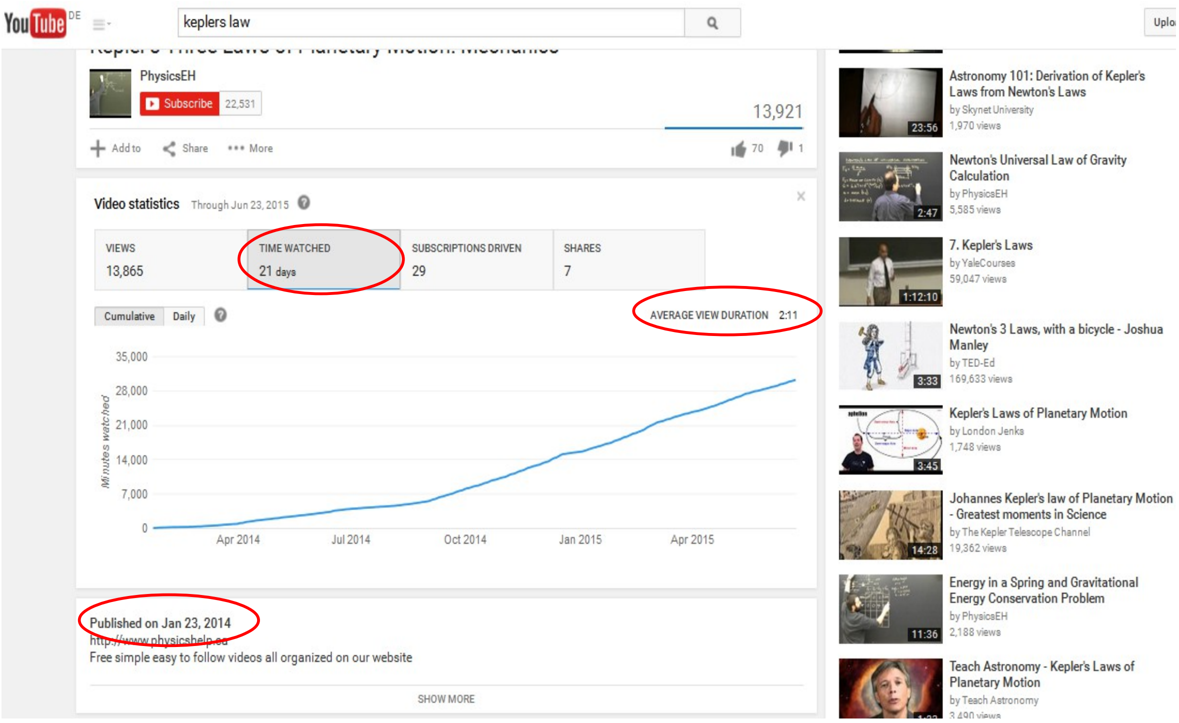


Figure 4: The *Average View Duration* can be found and is presented as shown above. Here the average view duration of video **EP04p** as of 08/June/2015 can be seen.

- **Comments**

The **Comments** provide by far the most intense communication channel between explainer and addressees since the explainer can reply directly to an addressee's comment. They can be found right below the video and are considered an engagement metric by Google Developers (2015), underlining the importance of their status. Their presentation can be seen in Figure 5.

Here, new topics are being suggested and constructive criticisms given by anyone with an account who likes to give his or her comment on the video. Sometimes a whole dialogue can be found, discussing the topic, improving the explanation or answering questions. (For a detailed description on **Comments** see 3.4.1). On the other hand,

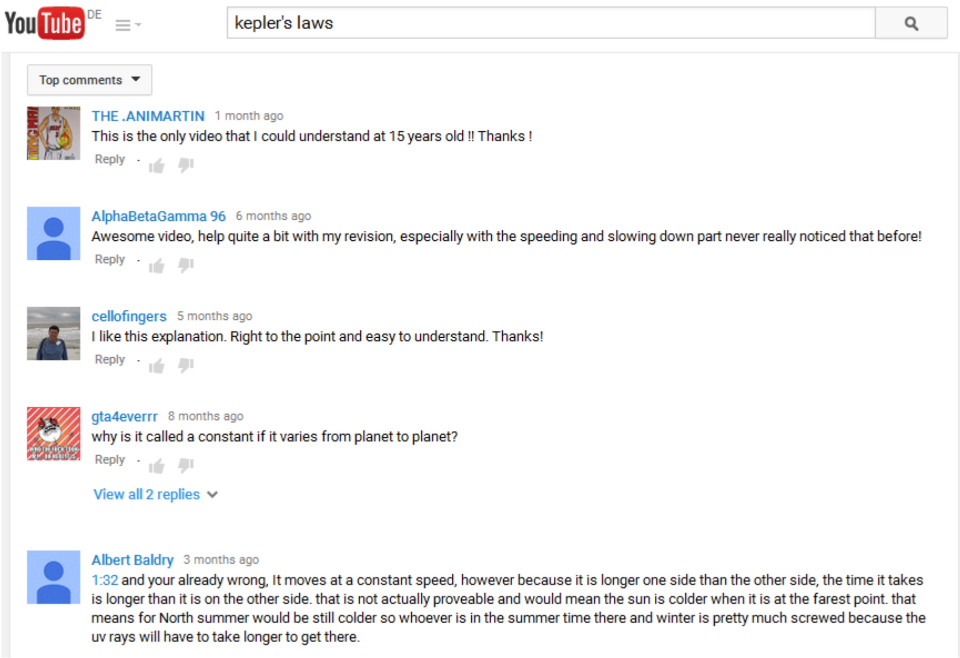


Figure 5: Comments given to video **EP04p** as of 08/June/2015.

Comments also accumulate a lot of trolls which *The Guardian* describes as "*the act of posting disruptive or inflammatory comments online in order to provoke fellow readers*" (The Guardian, 2014) and hurtful comments. Their psychological effect is still unknown today (The Guardian, 2014).

Knowing YouTube's data collection methods, **Views** might be considered less quality-related than **Likes**, which itself might be even less related to explaining quality than **Average View Duration** or **Comments**. Furthermore, Stangl (2015) states that outer criteria are usually questionable as their qualities are themselves questionable. Thus, they are considered carefully and with caution within this work.

A complete list of all short-listed videos, their URLs and all other surface feature data can be found in Chapter 7.

- **Categories**

The process of attaining the 45 categories used in this work to gauge explanatory quality is briefly explained not only for gaining insight into the process itself, but also to honour the work of the people who developed these categories.

In 2013, Kulgemeyer and Schecker (2013) developed a method to qualitatively assess science communication competence (SCC) by videotaping 46 high school students in a peer-to-peer tutoring environment. The researchers' motivation was a lack of theoretical background or guidelines for SCC and its assessment. Their main task therefore was to link abstract educational objectives with concrete problems such as physics didactics-specific competences with students' explaining skills. Firstly, they had constructed a normative competence model, which had then been revised due to the evaluation of students' performance which added descriptive items to the model. The model is based on both addressee-adequacy, domain-orientation as well as seeing transfer of knowledge as constructivist according to Rusch's and Merten's communication models, see 2.1. Secondly, they analysed the explainers' reactions to prompts from trained addressees asking to simplify their explanation by qualitative content analyse. These reactions were allotted to categories and after further refining, they came up with 16 categories which, they found out, do overlap with some elements from both Wellenreuther's and Brown's findings. They determined the most important SCC categories to be:

1. Prepare an introduction
2. Produce and use graphs
3. Direct address or call for action
4. Confirm understanding
5. Concise answers
6. Give examples

All 16 SCC categories are listed in the annex and form the basis for Kulgemeyer and Tomczyszyn's further research.

With those previous 16 categories not encompassing gestures, writing, animation or experiments, Kulgemeyer and Tomczyszyn decided to incorporate those supporting elements into the model. Consequently, the focus has been shifted from the oral explanation itself to the **process** of giving an explanation (Tomczyszyn, Kulgemeyer & Schecker, 2014). Their objective was to model the skill to deliver a rich and varied explanation and to validate that model in an activity-oriented situation. For this reason, they videotaped and analysed 26 highly standardized expert-novice-dialogues in a pilot study qualitatively. That dialogues are to simulate a real learning situation that could be found in students' daily routine. The standardization covers topics, preparation length, documents, aids and a constant dialogue duration of ten minutes. They used prompts from *trained novices* who asked for, say, a numerical example or use of less technical terms. Subsequently, they categorized the explainers' reactions from their pilot study from an addressee-adequacy and domain-specific point of view as well as complementing their categories for results found in previous studies. Their 26 categories achieved a Fleiss's kappa of $\kappa = 0,798$ and a reliability of Cronbach's $\alpha = 0.735$.

After the pilot study, Kulgemeyer and Tomczyszyn (in press) further improved the measure by adding *"[c]ategories and indicators for the quality of explaining [that] were derived inductively from the reactions to [the] prompts and deductively from the theory of explaining"* (p.2). The theory includes constructivism and communication with focus on the addressee and his or her needs as outlined in Chapter 2.1. They also considered the six main SCC categories outlined above and added four that also appeared most frequently. These four include (Kulgemeyer & Tomczyszyn, in press, p. 9, translation by the author):

1. Consider addressee's prior knowledge
2. Emphasise important points
3. Give opportunities for being asked
4. Logical reference between content and explanation

Equally, the communication model developed by Kulgemeyer and Schecker around 2009 was included (see 2.1) after having slightly changed the variables from factual content aspects, representation form, context and code into **mathematization, code, examples & analogies** and **graphic presentations** which includes both pictures and objects. The variable mathematization is important as it includes "the...the"-relations and different representations such as graphs and symbols like algebraic structures. To find categories, they applied qualitative content analysis and compared those to the ten SCC categories and the variables from the communication model. The whole process is

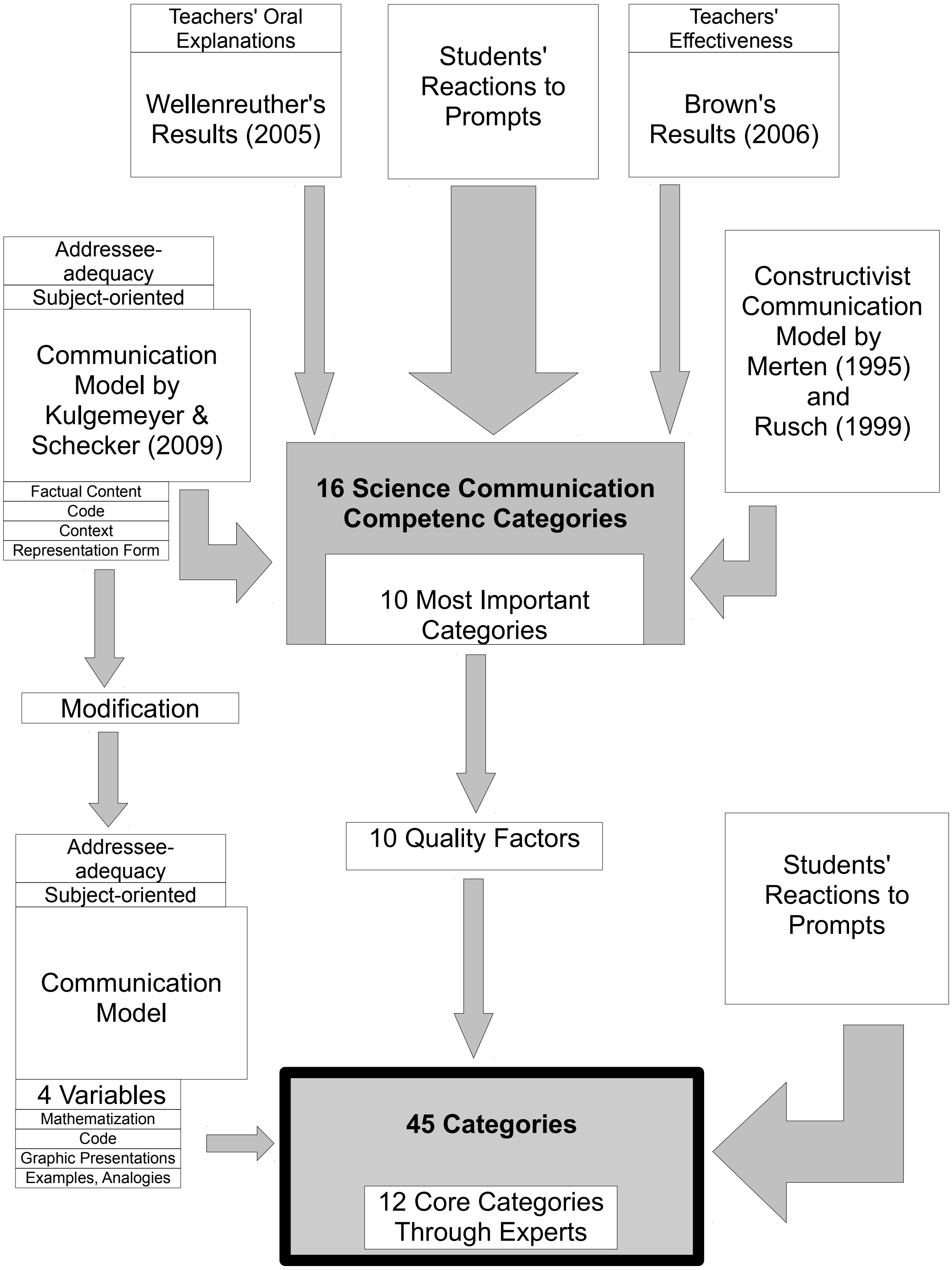


Figure 6: Simplified schematics of the process to attain 45 categories that make up a good explanation (Kulgemeyer & Schecker, 2013; Kulgemeyer & Tomczyszyn, in press). The 12 core categories are listed in the annex.

illustrated in Figure 6. After having determined 45 categories that fulfil the criteria, a measure was developed to give quantitative data and the measure was verified by concurrent validity (experts), content validity (encompassing all variables from the model) and reliability measures, which the measure satisfies as both reliability and validity reach a value of more than 0.7 each.

The twelve core categories were derived through ratings conducted by experts *"since a valid interpretation of explaining skill is reflected insofar as it can predict experts' decision on two explanations with* $\kappa = 0.75$*. (It proves also reliable on different topics with Cronbach's* $\alpha = 0.772$*)"* (Kulgemeyer & Tomczyszyn, in press, p. 23). The dozen core categories can be found in the annex.

For a more detailed list with all 45 categories including examples for each category, see annex (Chapter 10).

3.3.2 Data for Predictive Validity

The data for research question (3), which investigates the learning impact of the explanatory videos, are collected in a field study by comparing scores from two knowledge tests - one taken before (pre-test) and one after (post-test) an explanatory video on Kepler's Laws has been watched. These knowledge tests are exactly the same, helping participants to focus on the topic since they already know the test questions of the post-test before watching the video. The participants are allowed to take notes as the knowledge test shall not measure memory capacity. The post-test additionally contains a brief questionnaire asking the participants whether he or she has ever used explanatory videos, among others. The score is derived by summing up one point for each question answered correctly, while questions answered incorrectly are counted as naught. The two scores can then be compared by their means. This procedure is done with two unique videos in order to compare the learning impact of either explanatory videos.

- **Test Development**

The basic idea is of this data collection is that participants fill out a pre-test on Kepler's Laws, then they watch a video that deals with the explanation of the three laws. This

Table 2: The main categories and their subcategories. Subcategories with an "] " denote a negative category. (This list is based on Tomczyszyn's, see (Chapter 10))

Main Category	**Subcategory**
① Content	wrong topic]
	mistake]
	mistake corrected
② Medium - Verbal: Structure	outlook
	review
	summary
	encouragement
	patronizing comment]
	emphasizing important points
	justifying approach
	addressing pre-and misconceptions
③ Medium - Verbal: Lecture	anthropomorphism
	paraphrasing technical terms
	comment technical term close to everyday life
	comment technical term close to subject
	uncomment technical term]
④ Medium - Verbal: Context	including addressee
	example close to everyday life
	abstract example
	inadequate example
	without context
	connecting examples
	connecting contexts
⑤ Mathematization	numerical example
	formulae
	"the...the"-relation
	mathematical terms & idealisations
⑥ Medium - Verbal: Interrogation	prognosis
	diagnostically affirm comprehension
	inquire previous experience
	directly test previous knowledge
	diagnostically test previous knowledge
	directly affirm comprehension
	further questions
⑦ Call for Action	call for action
⑧ Medium - Non-Verbal	real figure
	analogical figure
	logical figure
	objective aids: experimenting
	connecting objects with scenario
	writings
	connecting non-verbal elements
	create/amend figures
⑨ Medium - Non-Verbal: Structuring	interruptions]
	inadequate reply]

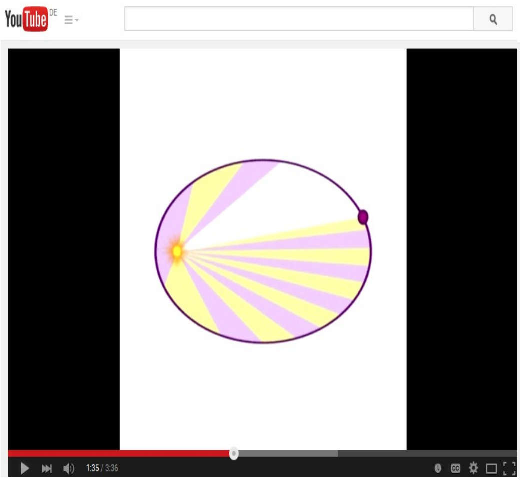
is then immediately followed by the same pre-test which incorporates a short questionnaire on their attitude towards explanatory videos. The objective of this process is to measure the learning impact caused by the explanation used in the video on each student individually.

The test itself is a brief performance paper-and-pencil test measuring knowledge of facts, therefore it was decided to use a multiple-choice format with only one out of four possible answers per question being correct in a pre-post-test design. Also, to allow comparison between the videos, only one test was developed and used for different videos. A code is added on the title page to enable pairing of the pre- and post-tests after execution. To develop the required test items for such a test, videos that come into consideration for the field study were being watched again and possible questions listed. Only videos in German were selected due to the location of the field study at University of Bremen, Germany. Then only those questions were selected that occurred in the majority of videos. After that, the questions were drafted and having chosen a tied answering format, possible answers were devised, too. It was ensured that the test items were drafted value-free and in a neutral language. For example questions, see Chapter 10. The language of the items is also clear but domain-specific, therefore it is peppered with technical terms but still appropriate for major and minor physics students. Each item contains only one question and is free of double negatives.

Two videos were chosen that are equal in run-time (approximately four minutes) but as far apart as possible according to awarded category points (CPs) to allow a clear rank (see 3.4.1). (For learning on how the CPs are awarded, see 3.4.1.) Since only four videos remain at both 6 CPs and 9 CPs, it was decided on video **DT04** (6 CP) and **DU20** (9 CP) and subsequently downloaded from YouTube to make them available offline. Video **DU20** was cut to under four minutes, since Kepler's Laws is only part of the first few minutes, after that the explainer describes an app and its handling. This also ensures equal run-time of the videos used in the study. An expert also agreed with the superiority of video **DU20** to **DT04**, reflected in their respective CP points. For more details on the two videos, see Figure 7.

A score could then be awarded to each student's performance before and after having watched the video. It was calculated by assigning one point for each question answered

T-Video: DT04
(URL: See Chapter 7)



(Sub)categories used:

paraphrasing technical terms

comment technical term close to subject

abstract example

"the...the"-relation

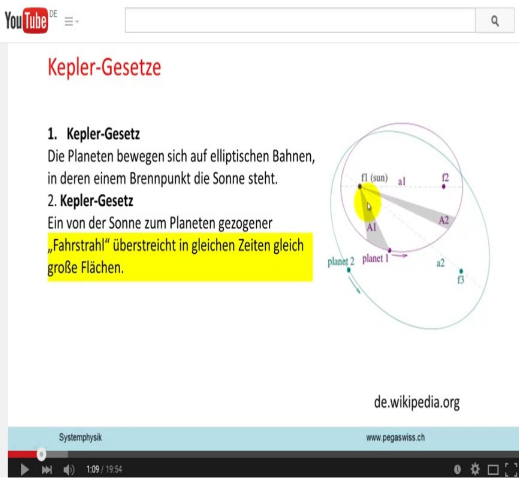
mathematical terms / idealisation

analogical figure

---

6 CPs

U-Video: DU20
(URL: See Chapter 7)



(Sub)categories used:

summary

emphasizing important points

justifying approach

paraphrasing technical terms

comment technical term close to subject

uncomment technical terms] (-1)

example close to everyday life

formulae

mathematical terms / idealisation

logical figure

writing

---

10 CPs – 1 CP = **9 CPs**

Figure 7: The two videos selected for the field study. A typical scene from video DT04 (LEFT) and DU20 (RIGHT) as well as the subcategories that make up the CPs. Note that DU20 also includes a subcategory that must be subtracted from the CP value.

correctly and naught for incorrect answers. Later, the two learning impacts of the two videos can be ranked and compared to the rank derived by the measure.

The questionnaire attached to the post-test uses an odd but symmetrical, balanced and equidistant Likert-scale to be ticked off with an average of three implying that participants neither agree nor disagree with the item. The neutral centre is recommended by Stangl (2015) when measuring interest, which this test does, to allow a neutral position. The scale uses language to describe its value rather than numbers. A mean of four, for instance, suggests that they agreed a little with the statements in the items. The items in the questionnaire are more easily phrased than those in the test to prevent interpretation and assumptions. They should appear neutral and value-free.

The whole knowledge test and questionnaire can be found in the annex.

- **Test Execution**

After having finished the development, the knowledge test as well as the questionnaire underwent a pilot study with ten high school students, resulting in the elimination of three questions and thus ensuring that all questions can be answered after having watched the two chosen explanatory videos as well as improvement of the figures in item number three as some students were confused by one picture. Furthermore, the code was revised to reach a higher degree of diversity. The test also has one peculiarity: Item number one can only be answered by video **DU20** and item seven only by **DT04**, awarding both videos a maximum of six points, although the complete test has seven questions. The pilot study also shows that the execution takes about 20 to 25 minutes, which is considered an acceptable length not to exhaust or overwhelm the participants and to gain access to lectures and seminars needed for the field study at Bremen University. For time efficiency reasons, the test is conducted in groups of five to 20 participants, dependent on the attendance of the seminars. Due to the limitation of conducting the field study during class, a whole group is examined rather than individuals. Thus, making it impossible to simulate a YouTube user's original behaviour in manipulating the video using playback, pause or review.

The execution follows a strict procedure to allow objectivity and standardization by using a *briefing check list*, consisting of the following items:

- Reception
- Participation is voluntarily
- Anonymity is paramount
- Sequence (pre-test, video (shown once), post-test)
- Similarity of tests (post-test contains one additional page)
- Encourage taking notes and allowing their use during post-test
- Describe code and its purpose
- Emphasize: only one question per item is correct, hint: print is two-sided
- Duration: approx. 20 minutes
- The test is only collected when all participants have finished
- No assistance is provided
- No talking

After a short address of welcome, the procedure is explained as well as the duration by stressing the voluntariness and anonymity of the test. Participants are allowed to take notes since the purpose of the test is not to measure memory capacity, but the explanatory quality of the video. As all the information come from the video, they are also allowed to use them during the post-test. The decision to show the video only once aims at identifying the first impression of the video on someone's learning impact. Also, showing the video only once might increase the participants' attention and alertness. Another reason for this decision was that participants know that both tests are the same and therefore might scan the video for answers to the questions; showing them the video twice might measure their scanning competence rather than their dependence on the information given in the explanation. Having made the experience during the pilot study that some question the code, the purpose of the code is explained during the briefing, so is the fact that only one answer per item is correct.

For the execution, the following utensils are used: laptop computer, video **DU20** (shortened) and **DT04**, loudspeakers, projector, extension cords, briefing check list and copies of the pre- and post-test.

3.4 Data Evaluation

This section is about describing the data evaluation process, and in the case of the paper-and-pencil test, it also serves to guarantee the compliance with quality criteria.

3.4.1 Quantitative & Qualitative Content Analyse

This subsection presents both the method chosen to perform quantitative content analysis according to Mayring and the path to derive the categories needed to determine content-relevant comments via qualitative content analysis according to Mayring (2000).

- **Quantitative Content Analysis: The Measure**

Quantitative content analysis according to Mayring uses developed subcategories for each variation in the explainer's explanation to which the content of the material is sorted accordingly (Stangl, 2015). Its objective is to *"allow an analysis of a meaning in form of frequency"* (Stangle, 2015, p. 1, translation by the author) to make comparisons or reveal tendencies. To make such an analysis possible, quantitative content analysis always uses some elements of qualitative content analysis because the occurrences have to be interpreted (Mayring, 2000). That is the reason why a reliability examination and an objectivity measure are needed to maintain its comparability and to prove its practical relevance (Mayring, 2000; Stangl, 2015). Here, the measure variations of an explanation become the explanation methods, or explaining aids to assess quantitatively the explaining quality of 40 YouTube videos on Kepler's Laws. The issue of making use of those categories on Youtube videos is broached in 2.2 and is made possible as both the 45 categories and the explanatory videos are based on the same didactic cores: addressee-adequacy, considering the addressee's prior knowledge and understanding learning as a constructivist activity.

The measure itself is derived by cumulating each occurrence of a subcategory as defined in 3.3.1 by awarding one point (+1) to its use by the explainer. Each category is treated equally in this process and subsequent uses of the same category are not counted since repetitions of the same wording or repeated use of a similar explaining aid without any variation are not considered a rich and varied explanation (see 2.1). As a result, the following assumption applies: The better the explaining skill, the more diverse the explanation, the more categories used, the higher the CPs awarded. The same holds true for **Views, Comments, Likes** and **Average View Duration**. Some categories, however, are associated with a discount of points as described in 3.4.1. Hence a simplified frequency analysis is applied to count the emergence of a subcategory: Frequency analysis usually counts each occurrence of a category but in this case it is only counted once as it is only of concern **if** the subcategory emerges, not how often that category is used repeatedly, or when it occurs. This approach also excludes valence, intensity and contingency analysis (Stangl, 2015), which add assessment, judgement and clustering, respectively. The applied frequency analysis yet permits an explainer who uses a lot of different explaining aids, or subcategories, to accumulate plenty of points. However, some categories are considered prejudicial to learning and are therefore assigned a negative point (-1) for its occurrence as it is not considered beneficial to the explanation. These categories encompass **wrong topic, mistake, patronizing comment, uncomment technical term, interruption** and **inadequate reply**. They are also marked by the symbol "] " in the category list found at 3.3.1. Using the results of the quantitative content analysis, an ordinal performance number can be calculated for each video, abbreviated by **CP(s)**, which stands for category point(s).

$$\boxed{CP = \sum X_+ - \sum X_-}$$

X_+ stands for all positive and X_- for all negative categories, as noted by "] " in Table 2. The subtraction of one point due to the category *mistake* can be eliminated by correcting the error or fault. The use of subcategory *mistake corrected*, basically neutralizes the aforementioned category. All other subtractions cannot be cancelled out. The awarded CPs are then noted for each video and used for correlation analysis to other surface features of the same video as well as finding two videos with different CP values for the field study.

- **Qualitative Content Analysis: Comments**

These strengths are (1) fitting the material into a model of communication, (2) analysing the content step-by-step, (3) categorizing and (4) maintaining quality criteria (Mayring, 2000; Stangl, 2015). Also, the data can then be used for quantitative content analysis.

Qualitative content analysis tries to open up the methods used in quantitative content analysis by *"preserving the strengths of content analysis in communication science [...] to develop qualitative procedures [...] which are methodological controlled"* (Mayring, 2000, p. 6).

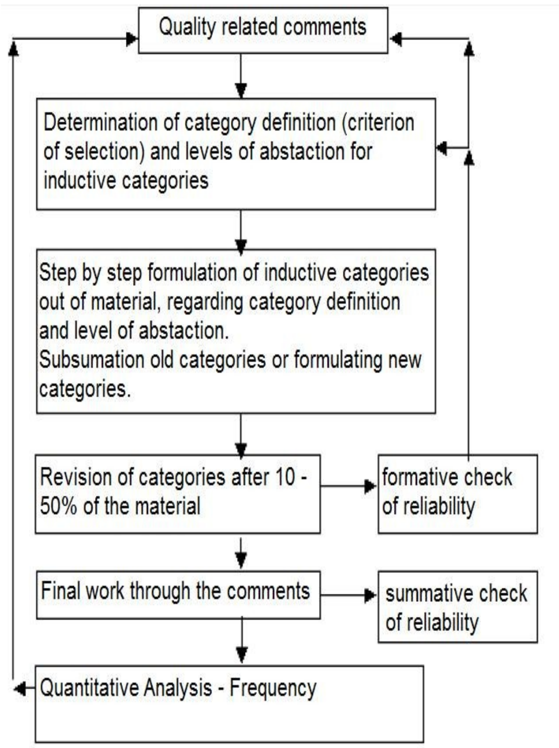


Figure 8: Mayring's suggested standardized path to develop categories inductively (Mayring, 2000, p. 4), topic adapted.

Stangl (2015) points out that, apart from finding the essence of data by means of qualitative content analysis, qualitative content analysis is a subjective method and as reliability quality criterion measures cannot be implemented directly, it becomes paramount to illustrate the way the categories have been developed. For this reason, Mayring suggests a way as illustrated in Figure 8. Hence, in order to find content-related, or relevant, **Comments**, which may be connected to explaining quality, inductive categorizing is used. This method of qualitative content analysis uses the **Comments** submitted by YouTube viewers to chart categories exclusively. The material is worked through by means of the standardized path and the feedback bend was used to adapt the categories, reduce or summarize them and the reliability can be checked simultaneously in this process (Mayring 2000, p. 4). See 4.3 for categories and 4.3.3 for amount of relevant **Comments** per video.

3.4.2 Spearman's Rank Correlation Coefficient

With CP, **Views, Comments, Likes** and **Average View Duration** being of an ordinal scale, Spearman's rank correlation coefficient is used to calculate correlations. Spearman's rank correlation coefficient can only be used if more than five cases are included and if the distance between ranks are equidistant. Its value lies between -1 and +1. Also, as several ranks are occupied by more than one video after having awarded each video its rank according to its CP points, Spearman's rank correlation coefficient with ties is applied. Moreover, Spearman's rank correlation coefficient does not require any distribution, but it only examines monotonically related data (Owen et al., 2015). It is calculated as follows:

$$\rho = \frac{\frac{1}{n}\sum_i (rg(x_i)rg(y_i)) - \overline{rg_x rg_y}}{\sigma_{rg_x}\sigma_{rg_y}}$$

, where n stand for the number of cases, $rg(x_i)$ denotes the rank of x_i, $\overline{rg_x rg_y}$ describes the product of the two means of the ranks and σ_{rg_x} stands for the standard deviation of the ranks.

To maintain objectivity, the calculation was performed by PSPP and Excel, respectively. Kendall's Tau was considered but later dropped as it uses the difference between ranks only instead of the distance between ranks. James Lani also note that *"[i]n most of the situations, the interpretations of Kendall's tau and Spearman's rank correlation coefficient are very similar and thus invariably lead to the same inferences.* (Lani et al., 2015, p.1). Furthermore, to determine significance, the following graph (Figure 9) as well as *Upper Critical Values of Spearman's Rank Correlation Coefficient* R_S found in the annex are used rather than the t-test, which requires metric data. The columns 0.025 was used which represents the 95 % confidence interval for a two-tailed distribution.

Spearman's rank correlation coefficient allows similar interpretations as Pearson's linear correlation coefficient (Owen et al., 2015), see Table 3.

Table 3: Interpretation of Spearman's ρ according to Owen (2015).

0.00 - 0.19	very weak
0.20 - 0.39	weak
0.40 - 0.59	moderate
0.60 - 0.79	strong
0.80 - 1.00	very strong

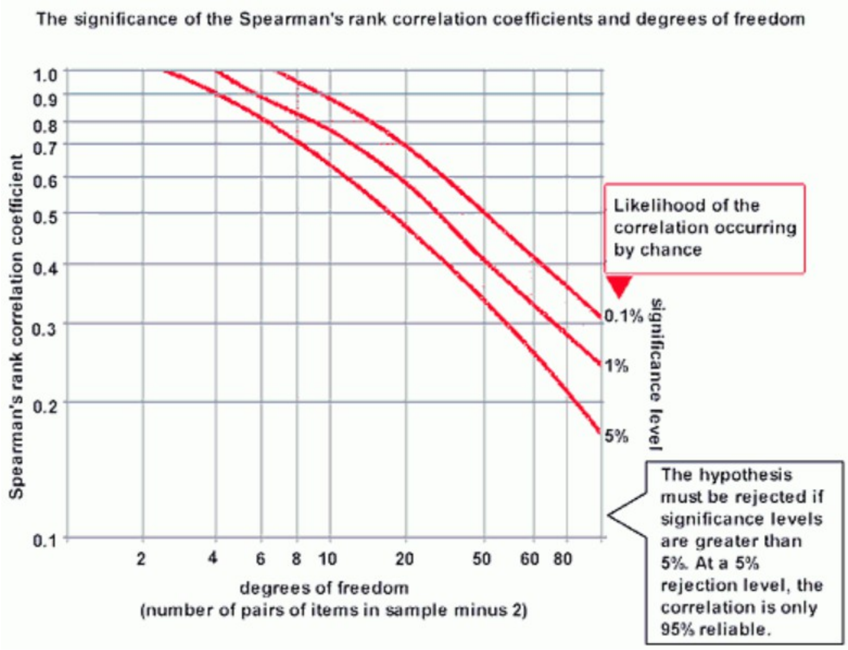


Figure 9: Significance of Spearman's Rank Correlation Coefficient (Barcelona Field Studies Centre, 2015).

3.4.3 Comparison of Test Means

Due to the complexity of measuring explaining skills, it was decided not to have students rank different explanatory videos, rather it was decided on a more cautious approach: to evaluate the effect of two videos on students' gain of knowledge. Thus, it allows a quantitative data evaluation method: Video **DU20** is ranked higher than **DT04** as per awarded CPs. The hypothesis now is whether video **DU20** has a higher learning impact than the other video, thus the null hypothesis is H_0: *video **DU20** does not have a higher learning impact than video **DT04***. This is measured by comparing means of achieved points per correctly answered questions in a knowledge test (see 3.3.2).

The test data are entered into a spreadsheet without any interpretations, using the transform capability of PSPP to interpret answers. Then the amount of correct answers is accumulated, once for the pre-test and once for the post-test, ranging from naught to six points, for each participant individually.

To determine means, firstly the data was tested on normal distribution via Kolmogorov-Smirnov test incorporated in PSPP to allow the application of a t-test. In this case the null hypothesis is $H_0 =$ *no difference between sample distribution and normal distribution*, using a two-tailed distribution and a confidence level of 95 %. Secondly, the means M as well as sample standard deviation SD is calculated by PSPP. Thirdly, the significance via a paired-samples t-test, and independent samples t-test respectively, for metric data and the effect sizes are determined. The effect size *Cohen's d* describes the effect whenever the difference of the means is certain: If the effect size is 0.2, it is considered a small effect, 0.5 is a moderate effect and 0.8 is considered a strong effect. Both the test and the effect size are calculated via the three formulae given below:

$$t(df) = \frac{M_{post} - M_{pre}}{\frac{SD}{\sqrt{n}}}$$

$$d = \frac{M_{pre} - M_{post}}{\sqrt{SD_{pre}^2 + SD_{post}^2}}$$

n stands for the sample size, df for degrees of freedom, SD_{pre} and SD_{post} for the standard deviation derived from the pre-test and post-test, respectively. Similarly, M_{pre} and M_{post} stand for the two means.

3.5 Quality Criteria

The quality of test instruments must be known to determine the accuracy and universal character of that instrument as well as to know the influences of chance on the test results. Moreover, it shall prevent subjective influences (Wellenreuther, 2000). Another reason to attain quality standards is to maintain comparability and achieve access of statistical evaluation procedures (Stangl, 2015). The main quality criteria are objectivity, reliability and validity (Wellenreuther, 2000).

The main elements to maintain **objectivity** affect execution, evaluation and interpretation, whereas **reliability** describes the outcomes of repetitive executions which should result in equal data. Thus, it employs parallel testing, re-tests and the inner consistency of a test (Wellenreuther, 2000; Stangl, 2015).

Validity is the most important of the quality criteria as it scrutinizes the objective of a test or its agreement with outer criteria (Wellenreuther, 2000). *"The better a measurement process maps a theoretical term, the higher is its validity."* (Wellenreuther, 2000, p. 273, translation by the author). Part of the research question (2) is to be answered by determining the concurrent validities of the measure described above. Concurrent validity examines correlations to data collected at the same time as those for the measure, whilst predictive validity examines correlation to data collected later and is supposed to predict certain outcomes, as investigated by research question (3). Both aforementioned validities are considered part of criterion validity (Stangl, 2015), in contrast to convergent and discriminant validities, which are counted to construct validities (Stangl, 2015), together with factor analysis. Convergent validity gauges agreement to other validated constructs it is supposed to correlate with, whereas discriminant validity verifies predicted disagreement (Stangl, 2015). The two latter mentioned validities are not part of this work because the construct has already been proven to measure what it is supposed to measure - explaining quality. Another important validity deals with the content and checks whether the content of a test or instrument truly covers all relevant aspects that belong to a certain field that is to be examined (Stangl, 2015).

Factor analysis, though, was not used as it would have required more short-listed videos. The minimal number of cases is stated to be 60 to run an informative factor analysis (Bühner, 2006).

3.5.1 Quality Criteria of the Measure

For independence of the measure from different raters Cohen's kappa is used to quantifying the objectivity of the measure, besides standardizing execution and evaluation tasks performed by software programs such as PSPP. Cronbach's alpha is applied to determine reliability of the measure as re-and parallel testing of the measure would be counter-productive because a re-test would basically be a revision (maybe useful for intra-rater correlation) and a parallel test would need a modified set of categories.

- **Objectivity: Cohen's kappa**

To show the independence of an instrument from the rater, the same data should be produced independently of whoever conducts the examination when using the same materials and conducting equal data evaluation methods (Wellenreuther, 2000). Thus, a value is needed to quantify the two raters' agreement and the impact of chance with *"the inter-rater reliability help[ing] create a degree of objectivity"* (Davis, 2015, p. 1). This is accomplished by calculating Cohen's kappa, which determines inter-rater agreement with multiple nominal scale variables, which, in this case, are the different categories that quantify a rich and varied explanation (see 3.3.1). As it appears unacceptable for both raters to examine all short-listed videos, approximately ten per cent of all videos are inter-rated.

The formulae are (Fleiss et al, 2003)

$$\kappa = \frac{p_0 - p_c}{1 - p_c}$$

, where p_0 is the total of agreements (found on the diagonal of the matrix) and p_c the total of agreements attributed to chance (found in the marginal totals) each in per cent and calculated as follow:

$$p_0 = \frac{\sum_{i=1}^{n} h_{ii}}{N}$$

$$p_c = \frac{\sum_{i=1}^{n} h_{i\cdot} h_{\cdot i}}{N^2}$$

All three formulae are based on the following (n x n)-matrix:

		Rater A				Totals
		Main Cat. ①	...	...	Main Cat. ⓝ	$\sum$
Rater B	Main Cat. ①	h_{11}	...	...	h_{1n}	$h_{1\cdot} = \sum_i^n h_{1i}$
	...	...	...	...	...	...
	...	...	...	...	...	...
	Main Cat. ⓝ	h_{n1}	...	...	h_{nn}	$h_{n\cdot} = \sum_i^n h_{ni}$
Totals	$\sum$	$h_{\cdot 1} = \sum_i^n h_{i1}$	...	...	$h_{\cdot n} = \sum_i^n h_{in}$	$\sum\sum = N$

The main categories ① to ⓝ can be found in 3.3. A Cohen's kappa value of more than 0.60 is considered good (Grouven et al., 2007; Bortz & Döring, 2006, cited by Hammann et al., 2014).

- **Reliability: Cronbach's alpha**

Reliability quantifies an instrument's stability and consistency with a number from zero to one, where one describes the highest precision level. A possible test reliable quantity is Cronbach's alpha which perceives each item of a test as an isolated test itself by examining correlation with the total score independently for each item. Hence, it measures the inner consistency of a test.

Cronbach's alpha is calculated by means of the following formula:

$$\alpha = \frac{n \cdot \bar{r}}{1 + (n-1)\bar{r}}$$

, where n stands for the number of items and $\bar{r}$ represents the average inter-correlation between items. Its calculation is performed by PSPP, to maintain evaluation objectivity. A prerequisite required for Cronbach's alpha is that all items possess the same variance. Cronbach's alpha can be interpreted as *"an estimate of the lowest accuracy of the test"* (Langer, 1999, p. 1, translation by the author). That means the more items are used for the total score, the higher the reliability. A value of 0.7 or more is usually considered reliable (Langer, 1999).

3.5.2 Quality Criteria of the Paper-and-Pencil Test

As only group means are compared, the quality criteria need be less strict because every test can be seen as a re-test. This is due to the fact that accuracy is determined by *"number of items times number of group members"* (Wellenreuther, 2000, p. 281). Moreover, students are allowed to take notes while watching the video to minimize the possibility to measure memory capacity.

- **Objectivity** The objectivity of the pencil-and-paper test is obtained by using a closed answering format (eliminating the use of open-answer interpretations), standardizing its execution by means of a check list to guarantee both equal conditions and completeness, and using software (PSPP) for data evaluation and interpretation. Also no interpretation of the data is done prior to entering the data into PSPP. For more information, see 3.3.2.

- **Reliability**
Reliability of the knowledge test is of minor interest, as the test only properly works in conjunction with the two selected videos, and due to the interest in group means only, every test can be seen as a re-test.

- **Validity**

Its validity is determined by an expert being familiar with the subject, the videos and explaining theories via its content validity.

4 Findings

This chapter lists the findings and interpretations not only of the correlations and test means, but also of the quality criteria. Their discussion and answers to the research questions can be found in Chapter 5.

4.1 Video Selection

Of the more than 100 unique videos some would be excluded from the short-listed ones for the following reasons: Firstly, videos that show recorded university lectures or school lessons, as they do not fall under the definition of an explanatory video, e.g. a video produced for science teaching explanations. On the other hand, videos will remain included if they sum up a lecture such as **EU09k** (reviewing is one use of explanatory videos), or if they compose a complete series of short explanatory videos such as **EP05ddd**. Secondly, videos will be excluded if they are longer than 1,000 seconds (in excess of 16 minutes), see Figure 10, unless they can be shortened by cutting out elements that do not refer to Kepler's Laws, e.g. when the explainer refers to and explains a computer application, which simulates the planets' motions. Hence, video **DU20**, used in the field test, was 19:54 minutes (1,194 seconds) long, but it could be shortened to 3:07 minutes (187 seconds).

The mean of the final selection of videos is M = 417.35 seconds (SD = 253.19, N = 37), and the data possess a mode and median of Mo = 201 seconds and Mdn = 338 seconds, respectively, compared to a previous mean of M = 488.7 seconds (SD = 355.02, N = 40), a median of Mdn = 361 seconds and an equal mode of Mo = 201 seconds.

Thus, the final sample consists of 37 videos on Kepler's Laws on Planetary Motion

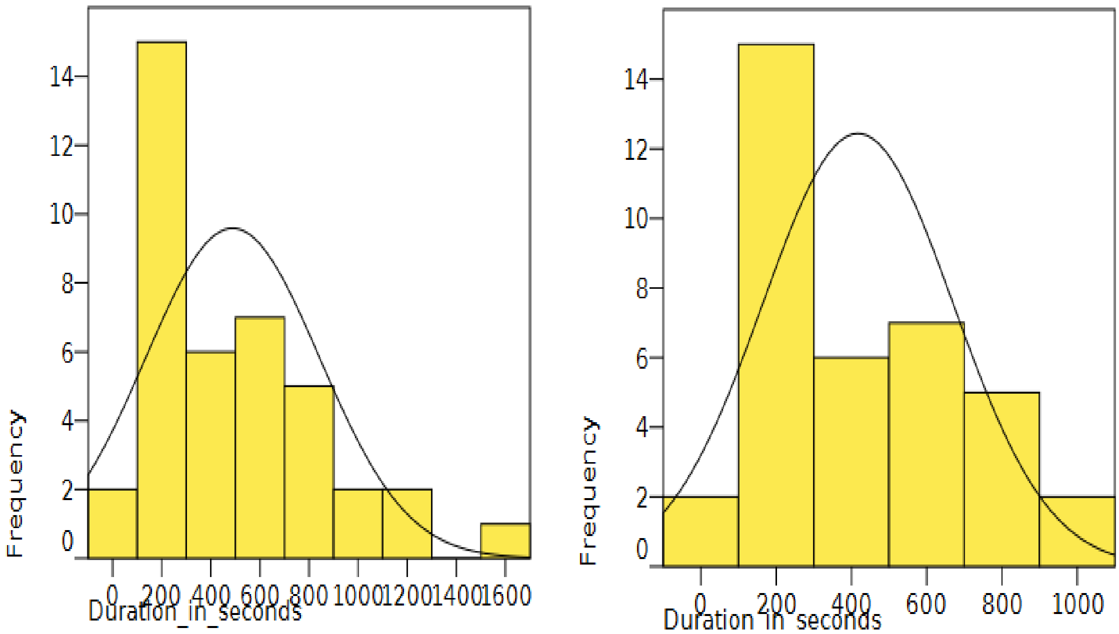


Figure 10: These histograms show the run-time distribution for all 40 short-listed videos (LEFT) including the positions of the three outlying videos with over 1000 seconds of run-time and the distribution of the selected 37 video (RIGHT). (PSPP)

in both German (five) and English (32), see Figure 11, of which 16 are made for private purpose (P-videos), 15 for educational purpose (U-videos), and the remaining six were originally distributed by commercial entities but found their way into YouTube's database (T-videos).

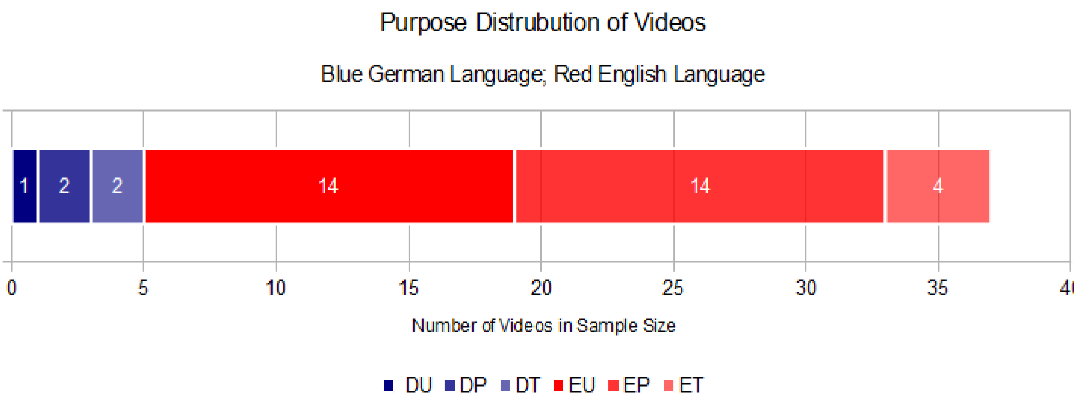


Figure 11: This bar chart shows the absolute frequency of all videos classified in language (German = D, English = E) and purpose (U = teaching, P = private, T = TV).

4.2 Quality Criteria

Considering the main differences between videos and expert-novice dialogues that were used to develop the categories the quality measures still hold true, which is supported by the filtered similarities within the underlying methods to give a rich and varied explanation as found in Chapter 2.1. The aforementioned differences include setting, preparation, video length, topic, communication channels -direct in the dialogue and asynchronous, if any, in the videos-, prompts, no time limits, more diverse and untrained addressees, and number of trials that can simply be cut off in postvideo productions.

4.2.1 Cohen's kappa

The Cohen's kappa is determined by using five out of 37 videos (approx. 13 % of all videos), including those videos that are part of the field study. These videos are **DU20** (shortened), **DU16, DP11, DP05**, and **DT04**. It is calculated by counting agreements and disagreements within a main category by accumulating agree- and disagreements in the subcategories. Rater B uses the same materials (detailed list of categories, see annex) and videos as rater A does.

A filled-out cell in the matrix on the following page (p. 50) denotes number of agreements within the main categories ① to ⑨, a cell within ⓧ indicates the number of disagreements in a main category whereas a blank cell represents the number zero. An agreement means either both raters select subcategory i, or both raters do not select subcategory i; otherwise it is seen as a disagreement

The percentage of agreements is $p_0 = 0.878$, and the number of chance agreement is calculated to be $p_c = 0.126$, resulting in a Cohen's kappa of $\underline{\kappa = 0.860}$, which can be considered very good (Altman, 1997). The measure reached a Fleiss' kappa of $\kappa = 0.750$ in previous applications (Kulgemeyer & Tomczyszyn, in press, p. 2), supporting the measure's independence of raters.

		Rater A										Totals
		①	②	③	④	⑤	⑥	⑦	⑧	⑨	ⓧ	$\sum$
Rater B	①	13										13
	②		33								3	36
	③			16							3	19
	④				30							30
	⑤					17						17
	⑥						35					35
	⑦							5				5
	⑧								43			43
	⑨									10		10
	ⓧ	2	4	4	5	3			4		0	22
Totals	$\sum$	15	37	20	35	20	35	5	47	10	6	230

4.2.2 Cronbach's alpha

Using all 45 categories, of which 31 occurred in all 37 videos cumulatively, the measure reaches a Cronbach's alpha of $\underline{\alpha = 0.690}$. With previous reliability determinations of the measure being $\alpha = 0.772$ in its original application on standardized 10-minute expert-novice dialogues (Kulgemeyer & Tomczyszyn, in press, p. 2), this result can be seen as satisfactory. For further discussion see Chapter 5.

4.3 General Results

As the CPs are ordinal data, box plots are used in to illustrate the distribution and to ease data interpretation.

4.3.1 Category Points (CPs)

Figure 12 shows the different box plots for different video classes.

The average median (of all videos, N = 37) is M = 8 CPs, ranging from two to 15, with 50 per cent of data being between six and eleven category points, data which is also averaged by videos for private purpose. Hence it appears that these 16 private videos

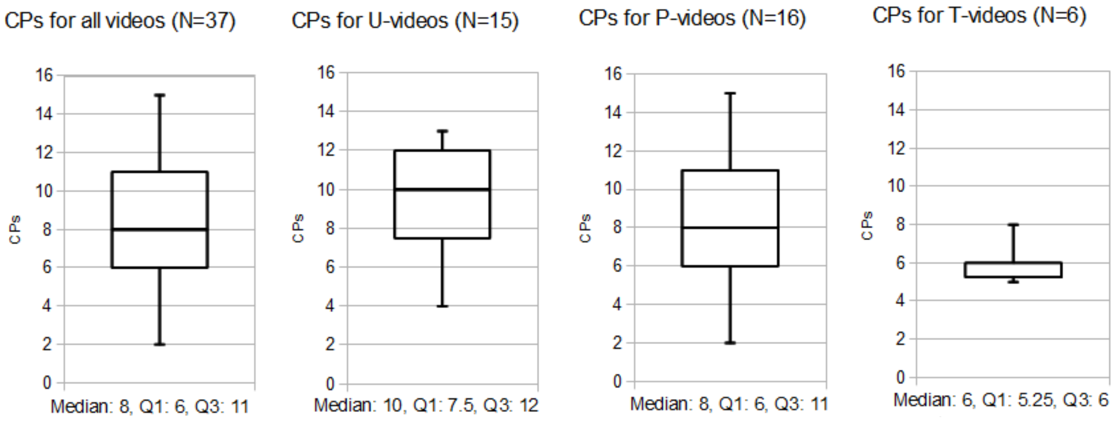


Figure 12: These box plots show the category point (CP) distribution for all 37 videos (LEFT), videos for teaching (U-videos), private (P-videos), and TV (T-videos) purposes (RIGHT).

(P-videos) represent the whole video sample of 37 videos. U-videos, teaching videos in university and classroom environments, average a slightly higher average median of Mdn = 10, and the CPs are closer together, accumulating 50 per cent of data between 7.5 and twelve CPs. This result can be expected when considering that U-videos are usually made by professional educators. Videos produced for TV and other mass disseminating media reach the lowest average median of six with the highest value being the average of P-videos and even lower than the median of U-videos. This might be due to the fact that T-videos are not solely explanatory videos as their purpose is to both illustrate and entertain at the same time.

4.3.2 Used Categories

The original 45 categories applied on 37 explanatory videos resulted in 31 subcategories to be assignable, leaving mainly categories out that deal with involving the addressee such as *Medium-Verbal: Interrogation, Call-For-Action*, and *Medium - Non-Verbal: Structuring* which describes interruptions that can be cut out in post-production of the videos. For a complete list of the remaining 37 subcategories, see Table 4.

Most categories used during an explanation are *Medium - Non-Verbal* and *Mathematization*, followed by *Medium - Verbal: Structure, Context*, and *Lecture*, separated by 18 total occurrences to the top two, as illustrated in Figure 13. On the other hand, categories such as *Call-For-Action* and *Medium - Verbal: Interrogation* do not occur, which was to be expected due to the limited interaction between explainer and addressee caused by the video format itself.

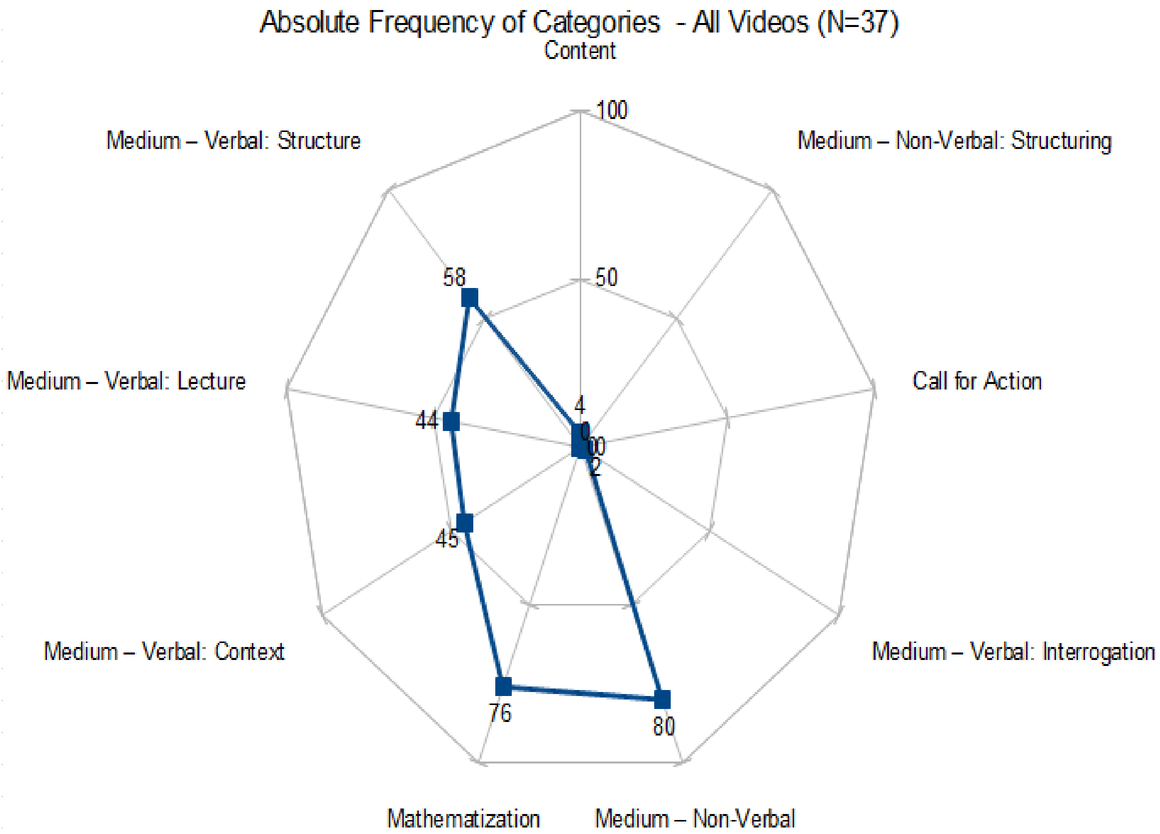


Figure 13: This net diagram shows the absolute frequency of occurrences of each main category for all 37 videos in nominal numbers.

After having analysed the categories, it can be said that all four communication variables (mathematization, code, examples & analogies, and graphic presentation) are still part of the 31 categories found. However, of ten quality factors (see 3.3) seven remain, as implementing *confirm understanding, direct address or call-for-action*, and *give opportunities for being asked* is restricted apart from commenting after having watched the video. But, given the limitation of interactions between the explainer and addressee, it is a satisfactory result that seven quality factors still remain incorporated. The missing three may explain, of course, the disappearance of some of the 14 categories that do not occur.

Table 4: The remaining 31 main categories and their subcategories for all 37 videos. The main categories retain their original numbering.

Main Category	**Subcategory**
① Content	mistake]
	mistake corrected
② Medium - Verbal: Structure	outlook
	review
	summary
	patronizing comment]
	emphasizing important points
	justifying approach
	addressing pre-and misconceptions
③ Medium - Verbal: Lecture	paraphrasing technical terms
	comment technical term close to everyday life
	comment technical term close to subject
	uncomment technical term]
④ Medium - Verbal: Context	including addressee
	example close to everyday life
	abstract example
	without context
	connecting examples
	connecting contexts
⑤ Mathematization	numerical example
	formulae
	”the...the”-relation
	mathematical terms & idealisations
⑦ Medium - Verbal: Interrogation	further questions
⑧ Medium - Non-Verbal	real figure
	analogical figure
	logical figure
	objective aids: experimenting
	connecting objects with scenario
	writings
	create/amend figures

Figure 14: These net diagrams show the absolute frequency of occurrences of each main category for the different purposes of videos in nominal values. U-videos TOP LEFT, P-videos TOP RIGHT ,and T-videos BOTTOM.

Taking a closer look at the main categories, separated by U,T-and P-videos, reveals that U-videos use *Mathematization* slightly more than *Medium-Non-Verbal* compared to all 37 cases (see Figures 13 and 14). Also, the use of different categories is more evenly distributed among various categories in U-videos. P-videos, on the other hand, use less of the category *Medium-Verbal: Lecture* and *Mathematization*. Moreover, the category *Medium -Non-Verbal* is the most used category in P-videos, compared to *Mathematization* in both U-videos and T-videos, which is unexpected. However, the data for T-videos must be approached with caution as only six videos make up this class of videos.

The most used subcategories of all videos are *mathematical terms & idealisation, formulae, logical figures, writing, abstract example, comment technical term closer to subject*, *summary, "the-the"-relation*, and *examples close to everyday life*. For more details, see Figure 15. It can also be seen that U-videos use a lot of verbal structuring elements such as *summary, emphasizing important points* as well as *justifying approach*. Similarly, U-videos, which top a dozen subcategories, use most subcategories when it comes to commenting like *paraphrasing, commenting technical terms close to subject*, and *comment close to everyday life* - a subcategory only found here. Moreover, mathematization categories can be found at U-videos as well such as *numerical example* and *"the...the"-relation*. P-videos, on the other hand, use most of *outlook* and *review* subcategories as well as illustrating examples. Most *examples close to everyday life* can be found here. Besides, P-video makers also may think that formulae sum up most of their explanations, as a result, the subcategory *formulae* can be found the most in P-videos. Subcategories of T-videos seem to cluster at mathematization, too, and they use the most analogical figures. Summaries seem to be an important feature of T-videos as well.

It is noteworthy that apart from minor mistakes that are being corrected during the run-time of the video, no video shows any mentionable mistakes. This might be due to the fact that Kepler's Laws themselves do not require mathematical derivations as they are findings based on observations. That is why the explanations mainly vary in their depth of mentioning implications and range of figures, mathematization, and structuring elements such as summaries and reviews which is also reflected in their CPs. In fact, that was one major criteria why the topic was chosen. In addition to that, directly

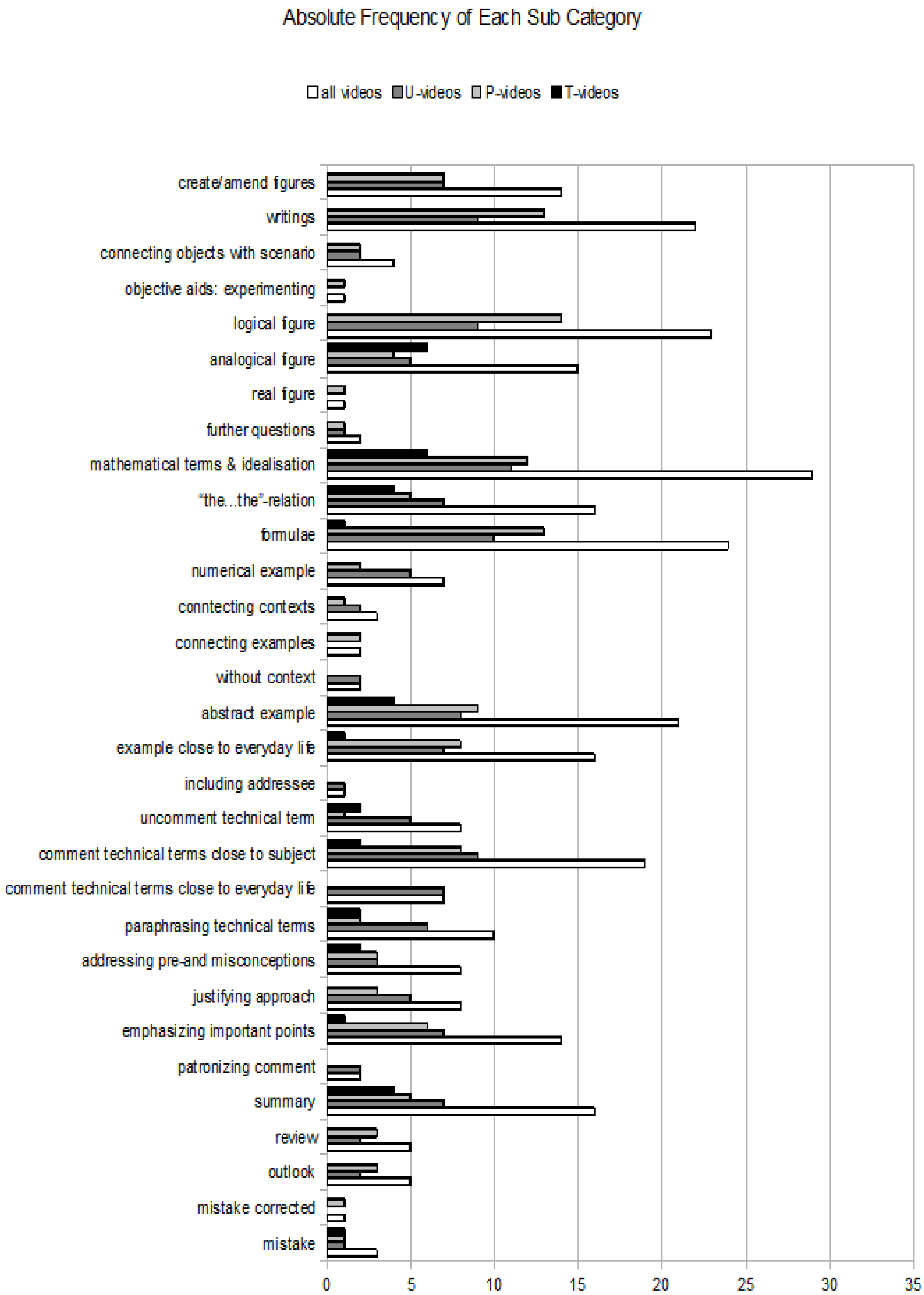


Figure 15: This graph shows all subcategories and their subsequent absolute frequency broken down into their occurrences within the different video classes.

addressing addressee's prior knowledge and misconceptions of Kepler's Laws as listed in 2.4 occur only eight times. It is, though, mostly found at other categories such as *emphasizing important points* in video **DT04**, e.g.: "Planets DO move on elliptical orbits, NOT circles" (capital letters symbolize stress), but in those cases it is not directly mentioned as a prior knowledge or misconception as required to label it as such.

All in all, no indications could be found not to use the measure on YouTube videos, as its results are consistent with expectations on, say, differences between U- and P-video as well as U-videos' slight superiority compared to P-and T-videos as well as with all four variables and seven quality factors remaining part of the measure. The question whether that use is also valid, is answered in 4.4 and 4.5.

4.3.3 Comments

Having used Mayring's qualitative content analysis, three categories of comments linked to quality criteria emerged. First, three categories, **Content**, **Explaining**, and **Explainer**, were defined. But after some videos another category was needed (**Used for**) and was confirmed having entered the feedback bend after approximately 15 videos. Then, a manual was compiled with category definition, subcategories and examples. The final work through the material proved the identified categories to be able to place all relevant comments into one of the four categories, see Table 5. The collection of comments took place on the second of May, 2015, as did all other data collection. All in all, 601 comments were analysed of which 169 were labelled as **Relevant Comments** and received an assignment to one of the categories mentioned above. The other comments include trolls, exclamations such as "thanks" or other comments not related in any way to the content or explanation. If more than one category may be considered for one comment, it is assigned to the category which the comment focusses on the most, see example in subcategory **Used For...** in Table 5. After reviewing all comments and assigning them to their category, the **Relevant Comments** per video are totalled to receive the value needed for correlation calculations. For reference, all totalled **Relevant Comments** per video are listed in Chapter 7.

Table 5: Inductive Categories for Relevant Comments.

Category	Subcategory	Example (copied directly from YouTube)
Comment on Content	Comprehension Question	"But what do you suggest the other point represents since there's only one sun in our Solar system. Don't get me wrong I completely understand the law but i just can't wrap my head around the relation of the second point to our solar system. I'd appreciate if you could clarify that for me thanks." ■ "does the moon obet kepler's seconds law?"
	Comment on Notation	"I have a question: For centripetal force, you wrote Fc, for centripetal acceleration, you wrote acp, so is it c or cp???"
	Further Question	"That was cool you have a soothing voice,can you please explain to me why the earth does not experience more solar heat when it is close to the sun. is it due to the motion of the sun in space." ■ "Is it really the Sun's center that's located at the focus? Or is it the barycenter of the planet-sun system?"
Comment on Explainer's Style	Good or Bad (incl. Reason)	"Thank you very much for posting this. Not only are your explanation of the concepts clear, but your narrations help us to understand the intuition that lead to these ideas." ■ "wow, am really amazed how compact and clear the information ,in 3min you learn more than you would learn in 3 hour ,before we used to study such law in 2or 3 classes by the time you reach the third law you already forgetting the first 2 ,but now the whole pic in front of you ,really amazing ,you always doing great job."
Comment on Explanation	Good or Bad (incl. Reason)	""I thought the first two laws were explained quite well...I thought the 3rd law was overly ellaborate and could have been explained alot easier."" ■ "You explained this very well. I watched a few others and just didn't get that third one, but you make it so clear. Thanks so much."
	More Videos to Come	"hey can you explain the other 2? please i really need" ■ "Haben Sie auch ein Video zur Gravitation?"
	Constructive Criticism	"Hi I am Astronomy student also in Virginia. I just have one critism of this Keppler hero. Fact is Comets go past earth repeatedly and therefore they have some type of orbit. So they are not unbound. They are bound same as the rest of us. Where is any unbound body? If it's a bound comet then it can not be positive energy. This is all just theory in my mind anyway. I do believe in the hyperbolic path in XYZ and the speed at different positions. Call them what you wish." ■ " Can you prove Kepler's 3rd law in the more general elliptical case? You have convinced me it is true for circular orbits, but what about elliptic ones?" ■ "You did not take the square root of the constants at the very end. The "some stuff" in the equation at the top should be 2pi/sqrt(Gm). This might confuse some people watching this video and learning these laws for the first tme"
Comment on Used For	Used for ...	"Wow great job! this video helped me SOOOOOOO much in writing my paper on Kepler's three laws! Very clear and understandable! Thank you really appreciate it" ■ "Currently learning about this at school, helpful ways to revise :)" ■ "Danke, dieser Film hat unserem Physik-Referat zu einer 1 verholfen"

4.4 Correlations

The main question is whether a measure based and developed on instant and synchronous communication ("prompts") can be used to determine explaining quality on videos without immediate but asynchronous interaction between explainer and addressee -if at all-, assuming the same basic didactic principles as evaluated in Chapter 2.1 hold true. Some hints to being founded on the same principles are portrayed in Chapter 2.2 as most categories are still occurring and differences in classes of videos are clearly noticeable. At this point, it is important to understand that it is not the objective to find the best explanation among 37 YouTube videos on Kepler's Laws as learning is way too complex a process and bound to an individual, but to distinguish between rich and varied explanations and those with fewer variations. Consequently, those with fewer variations in their explanation may be less suitable for a few more viewers as some learners' channels and needs may not be appealed to have a lasting learning impact on the viewer. This selective and constructivist view will warrant the labelling of the video with a more varied explanation as being superior. Having collected data on CPs and other surface data such as **Views, Average View Duration**, and **Likes** all collected at the same time (a prerequisite to concurrent validity), rank correlations is determined using Spearman's rank correlation coefficient since the data are of an ordinal scale. That is also the reason why Student's t-test, which requires metric data as it compares means, is not applied rather than *Upper Critical Values of Spearman's Rank Correlation Coefficient* R_S table (see annex) and the graph depicted in Figure 9.

Table 6: This table shows the Spearman's rank correlation coefficient of CPs to various surface features sorted by video class. Significance is denoted in bold as per *Upper Critical Values of Spearman's Rank Correlation Coefficient* R_S table (see annex) and Figure 8. The number of cases is given in parenthesis. A blank indicates that less than five cases are available. (PSPP)

Surface Feature	All Videos	U-Videos	P-videos	T-Videos
Views	0.14 (N=37)	0.38 (n=15)	0.29 (n=16)	-0.37 (n=6)
Views per month	0.14 (N=37)	0.31 (n=15)	0.17 (n=16)	-0.68 (n=6)
Average View Duration	0.55 (n=13)		0.66 (n=7)	
A.V.D. in %	0.09 (n=13)		0.38 (n=7)	
Likes	0.25 (N=37)	0.32 (n=15)	0.37 (n=16)	-0.44 (n=6)
Relevant Comments	**0.42 (N=37)**	0.49 (n=15)	**0.52 (n=16)**	-0.27 (n=6)
R.C. in %	**0.43 (n=29)**	0.48 (n=13)	0.27 (n=11)	-0.67 (n=5)
12 Core Categories	**0.72 (N=37)**	**0.69 (n=15)**	**0.81 (n=16)**	-0.49 (n=6)

Additionally, the convergent validity to twelve validated core categories (see 3.3 and Table 6) is determined to further quantify the construct validity of the measure. As the videos offer a great range of settings and run-times as well as quantifying explaining skill is still in its infancy, mainly tendencies can be observed.

The correlations show that a moderate but significant concurrent correlation can be determined for **Relevant Comments** both in total and in per cent as they are derived by equally focusing on explanation quality, whereas **Likes**, and **Views**, for instance, might possess other factors that influence their values more than quality measures, such as background designing, speaker's likeability and presentation form. Also, those data were pre-cast by YouTube without evaluating their data production, thus their quality criteria remain questionable (see 3.3). It is interesting to note that the moderate correlation between CPs and **Relevant Comments** increases for P-videos and disappears for U-videos. One reason for this result might be that viewers are more outspoken and willing to comment in a private environment as they do not have to fear repercussions from their educator who might have uploaded a video and requests his or her students to watch it. The moderate correlation further disappears when examining **Relevant Comments in %**, indicating that the total number of **Relevant Comments** might be a better indicator than relative frequency. However, note that the number of videos of the former is less than the latter because some videos have no comments which resulted in zero per cent.

On the other hand, the strong rank correlation to the twelve core categories is an unexpected result as most of the twelve categories do not appear within this video selection due to their focus on people's interaction. Yet, they still result in almost the same ranks. Private purpose videos even show a very strong rank correlation. This result supports the assumption that those twelve categories are the core of the previous 45 categories, or at least, they would have resulted in very similar ranks.

4.5 Paper-and-Pencil Test Results

The purpose of predictive validity is to predict certain outcomes due to prior measures: In this case explanatory video **DU20** receives nine CPs while video **DT04** is awarded six points, thus the former video is supposed to also have a slightly higher learning impact on students who have watched the video, than on those who have watched to latter.

4.5.1 Comparing Learning Impacts

The hypothesis here is whether a video determined to have the richer and more varied explanation compared to a video with a less rich and varied explanation due to their awarded CPs results in a better learning impact on students when they watch the video once; thus the null hypothesis H_0 is *the superior video does not have a higher learning impact than the inferior video*. However, learning impacts are dependent on many a variable like context and situation, making it arduous to determine their effect. That is why only tendencies might be observed.

All in all, a total of 27 participants took part in the field study, which took place at four different seminars at the University of Bremen from June 29^{th} to July 10^{th} in 2015. The students are almost evenly distributed with **DU20** having 14 and **DT04** 13 viewers. All students attend physics lectures as a minor subject in their second semester with their major being either chemistry or geology. On the one hand, this sample, though, cannot be seen as a representative sample of YouTube viewers for it only includes university students (from one city), which make up only one part of viewers. The whole population may include pupils, laymen, hobby astronomers, those interested in physics and so on. On the other hand, students (and pupils) may still make up the majority of science explanatory video viewers on YouTube.

The purpose is to simulate watching a YouTube science explanatory video similarly to the way the students would do at home, but to allow both subsequent comparing of the data and time efficiency, a standardized execution is selected. As a result, the video is shown to a whole group of students that could not interact with the video directly. Thus, it is not their personal and individual interaction with the video content and its resulting learning impact that is measured rather than the learning impact after having watched the video once without manipulating the playback.

First, the Kolmogorov-Smirnov goodness for normal distribution and metric data shows that all data are normally distributed as the null hypothesis (H_0 = *no difference between normal distribution and sample distribution*) cannot be rejected on a confidence level of 95 %, which allows the application of the t-test. See Figure 16 for their respective probability density functions.

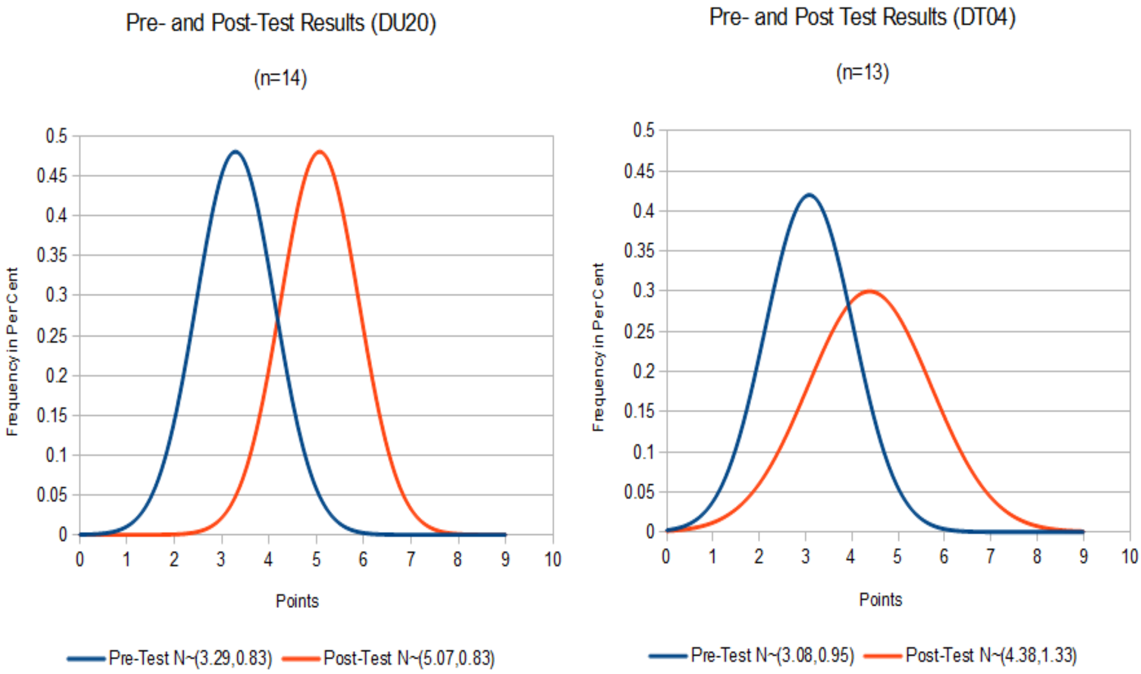


Figure 16: The test score results for video **DU20** are on the LEFT and for video **DT04** on the RIGHT.

The four means of the knowledge test score are:

- **DU20** - Pre-Test: M = 3.29 (SD = 0.83, n = 14), Mo = Mdn = 3.00
- **DU20** - Post-Test: M = 5.07 (SD = 0.83, n = 14), Mo = Mdn = 5.00
- **DT04** - Pre-Test: M = 3.08 (SD = 0.95, n = 13), Mo = Mdn = 3.00
- **DT04** - Post-Test: M = 4.38 (SD = 1.33, n = 13), Mo = Mdn = 5.00

The difference between these paired pre- and post-test means for each video is statistically significant (H_0 = *no difference between pre- and post-test means*):

- **DU20**: t(df = 13) = 7.49, p = 0.000
- **DT04**: t(df = 12) = 3.77, p = 0.003

These results indicate that both videos have a significant learning impact on students' knowledge. The test score of the post-test from video **DU20** is also slightly higher than the post-test score from video **DT04** and the standard deviation is smaller as well. Equally, it can be noted that the difference between pre-and post-test means from **DU20** is slightly higher (ΔM = 1.79) than from video **DT04** (ΔM = 1.31). The effect sizes of **DU20** and **DT04** are d_{DU20} = 1.52 and d_{DT04} = 0.80, respectively, indicating a strong and moderate effect in each case.

In order to determine whether the aforementioned differences between pre- and post-test score gains for both video is significant, the differences for each participant is calculated. Again, the data is normally distributed, resulting in N(1.79, 0.89) for **DU20** and N(1.31, 1.25) for **DT04**. All modes and medians are equal in value of Mo = Mdn = 2.00. Figure 17 illustrates their respective probability density function.

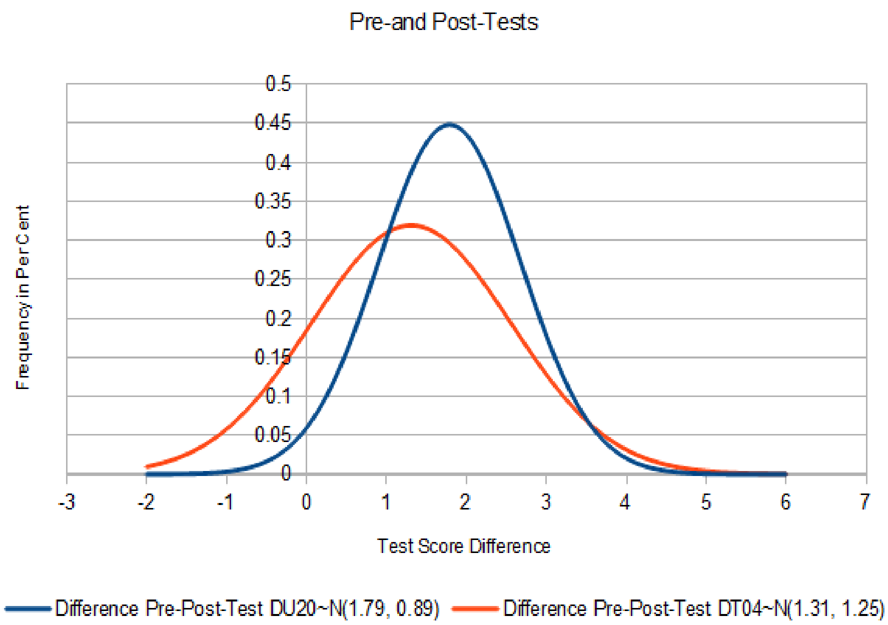


Figure 17: The differences of pre- and post test score means are not significant.

As the differences in test scores are also normally distributed, the independent-samples t-test is applied. The variance test (H_0 = *variances are equal*) shows equal variances assumed (p = 0.116), hence the t-test yields:

$$t(df = 25) = 1.15, p = 0.261, ns$$

The difference between pre- and post-test score gains from both videos is not significant. This may be due to the sample size of 27 participants, although a tendency that video **DU20** might have the higher learning impact than video **DT04** is clearly in the offing.

4.5.2 Students' Attitude Towards Explanatory Videos

The last five additional items of the post-test (see annex) are included to find out whether the participants have ever used science explanatory videos before as well as their attitudes towards those videos. Also, the purpose of the last item is to get to know students' media skills by asking them to gauge the explaining quality of video.

Item numbered as 12), which seeks to find out whether the participants have already used explanatory videos before, reaches a median of M = 3.00 (SD = 0.92), which indicates that most students sometimes use explanatory videos and is exactly centred at neutrality. Yet, being exactly centred at 3.00, it also indicates that an equal amount of students use explanatory videos both regularly and rarely.

Item numbered 11) and 15) inquires how much the participants like the video in general and how they would estimate the explaining quality of the video they were shown. Video **DT04** reached a mean of M = 3.84 (SD = 0.38) and M = 3.69 (SD = 0.48), whereas **DU20** received a median of M = 3.07 (SD = 0.62) and M = 3.07 (SD = 0.73) respectively. That means video **DT04** is generally more liked than **DU20**, and the participants also estimate the explaining quality to be better than of **DU20**. Even though students may like **DT04** better than **DU20**, the learning impact of video **DU20** is suggested to be slightly higher than that of video **DT04**. This might reveal a skewness between learning impact on the one hand and likeability as well as quality estimation on the other hand.

Given the above suggested skewness, it comes as no surprise that video **DT04** might be in favour of both preparing exams, tests as well as talks (item 13)), and encouraging the viewers to further engage with Kepler's Laws (item 14)). **DT04** scored M = 3.31 (SD = 0.95) at item 13) and M = 3.23 (SD = 1.01) at item 14), whereas **DU20** gained M = 2.93 (SD = 0.92) and M = 2.57 (SD = 0.94), respectively. Taking a closer at the data, another significant correlation emerged: Participants who like video **DT04** (item 11)) are likely to consider that video strongly when preparing exams, test and talks ($\rho = 0.61$, n = 13, t(df = 11) = 2.57). No such correlation, or any other, could be found for video **DU20** among those attitude questions.

Concluding, students use explanatory video from time to time, yet they might let themselves guide by other than explanatory quality criteria. Their quality assessment of the explanation may be based on non or low quality-related video features such as animations, colourful backgrounds, fast scene changes, and music rather than criteria that characterize the explanation itself. Besides, **Likes** and **Views** can also be mentioned in this list, as they do not correlate with explanatory quality; see Chapter 4.4.

5 Discussion

The purpose of this chapter is to discuss the findings and to put them into perspective with the previous findings as presented in Chapter 2.1. The final part of the chapter gives an outlook into further possible questions and endeavours.

The objectivity of the measure reached a Cohen's kappa of $\kappa = 0.860$ which can be seen as a very good value and reproves its independence of various raters. Also, the reliability value of $\alpha = 0.69$ is seen as a good result considering that the categories were developed for set ten-minute dialogues and not time-varied monologues found in explanatory videos. Yet, that it reaches such a high value shows that this established measure for dialogic explaining quality can be used on YouTube videos reliably and thus supports its application on this medium.

The research questions are presented in a reverse order due to research question (1) combines the findings from research questions (2) and (3).

Research Question (2):
Are there significant correlations between the measure and quality-related criteria such as *Views, Likes* and *Comments*? (concurrent validity)

As seen in Chapter 3.3, not all surface features provided by YouTube can be considered *quality-related*. Therefore, **Views** might be considered less quality-related than **Likes**, which itself might be even less related to explaining quality than **Average View Duration** or **Comments**. In this perspective, it is not surprising that the measure's rank received from their CPs correlate moderately at a value of $\rho = 0.42$ with **Relevant Comments** which were derived by applying qualitative content analysis by filtering out relevant comments that relate to the explainers' explanations. All other surface features show no such correlation, reflection their lack of quality-relation. Besides, YouTube's surface features are to be considered with caution as Stangl (2015) puts concurrent data's accuracy from third parties into questions. In this case, YouTube's data collection method could not be controlled or evaluated, questioning its quality especially when considering YouTube's deliberate manipulation of the appearance of videos to boost advertisement sales. Furthermore, Welbourne and Grant (2015) found out that user-generated content and channels are more popular than professionally generated science channels (here comparable to T-Videos) due to short videos and/or videos provided by the same popular explainer through his or her channels, hence they say that for a video to attract plenty of views it must attract a huge audience and fulfil their demands. However, views and content correctness do not align with each other (Welbourne & Grant, 2015). That might be a reason why CPs (explaining quality) and views do not correlate significantly in this work either as seen in Table 6.

Additionally, learning something new and adding knowledge into anybody's individual construct of the world is a highly demanding task that requires cognitive activity, which cannot be linked to surface features like **Views**, **Likes, Dislikes** or **Average View Duration** of explanatory videos at this point. Some viewers, though, may develop a need to make contact with the producer and/or explainer of an explanatory video via YouTube's **Comment** interface in order to talk, discuss or state questions. That could mean that those who give content-relevant comments are more activated cognitively by this video than by other explanatory videos. Hence, videos that cumulate plenty of those **Relevant Comments** are more successful in catching viewers attention as these videos might use either a more stimulating explanation or the explanation delivered is considered as a starting point for further learning progress. The data suggest that comments to a private entity are higher correlated to explanatory quality ($\rho = 0.52$) than to a professional person ($0.49, ns$) given the privacy and less judgemental nature of their exchange. This cognitive activation might be the reason why **Likes**, **Views**, and **Average View Time** are not directly connected to explaining quality of the video's content which resulted in no significant correlation (see Table 6) due to the viewers trial and error technique in locating videos with an appropriate explanation while their views are being accumulated for each video watched for whatever brief moment.

Yet, the correlation between the measure's CPs and **Relevant Comments** can be seen as a result that points in a direction that justifies deeper research into viewers' comments and their relationship to the explainers' explaining skill as well as examining differences between comments for U-,P-and T-videos. Furthermore, choosing to deliberately switch to another explanatory video can be interpreted as selectively seeking the explanation most suitable to the viewer's individual learning needs which might reflect some elements from Merten's communication model and underlines YouTube's huge database as a prerequisite for its success (Wolf, in press) since viewers are offered a vast range of videos to choose from.

Whether to use surface features to help viewers or educators, who wish to select and pick suitable videos for their learning or classroom teachings, distinguish the quality of science explanatory videos, it can be summarized that the total number of content-related comments might be considered a moderate indicator on the explaining quality of the video. Of course, watching the whole video and using personal experience as educators is still a more viable option, but for laymen, who use their subjective first impressions, the number of **Relevant Comments** might be of assistance to assure their decision for certain explanatory videos a little less subjectively.

Research Question (3):

Does the measure predict the learning impact of explanatory videos? (predictive validity)

Considering the small sample size, the predictive validity of the measure cannot be seen as rebutted as there is strong evidence that the measure is indeed capable of predicting learning impacts. Four pieces of evidence could be identified: First, video **DU20** has reached a higher test score mean (M = 5.07) than **DT04** (M = 4.38) in the post-test as seen in Figure 16. Second, the standard deviation of video **DU20** is smaller (SD = 0.83) than of video **DT04** (SD = 1.33), which might hint to a higher consistency in students' learning. Third, the difference between pre- and post-test score means is higher from **DU20** (ΔM = 1.79) than from **DT04** (ΔM = 1.31), as depicted in Figure 17. And fourth, the standard deviation of the test score differences derived from pre- and post-test scores is also smaller from **DU20** (SD = 0.89) than from **DT04** (SD = 1.25), again this might indicate a more stable learning outcome. These results all point into the same direction: That **DU20** offers the more successful explanation when considering the learning impact on students compared to **DT04**. However, due to the small sample size, all results must be considered as preliminary as they need further verification.

To put it in a nutshell, the underlying communication model by Kulgemeyer and Schecker, which emerged from models by Merten and Rusch as well as Wellenreuther's and Brown's findings, respectively the 45-category catalogue by Tomczyszyn and Kulgemeyer, seem so far capable of describing science teaching explanations from YouTube videos via its 45 categories realistically and in its entirety, as the ranks of the two video tend to be reproducible by students' gain of knowledge. Yet, students' intercourse with YouTube videos appears to be arbitrary and of little connection to its actual explanation quality, as depicted by the skewness between likeability of a videos and its impact on students' knowledge, see Chapter 4.5.2. It therefore becomes obvious that educators' objective opinion and guidance is still required to assist students in their selection process and handling of explanatory videos. This downside has already been identified by Diekmann, Gleiche and Weiner (2010).

Main Research Question (1):
Can an established measure for dialogic explaining quality be used to determine the quality of explanatory videos?

The validity of the measure's 45 categories, respectively its 12 core categories, was successful insofar as no indications could be found not to use them on science explanatory videos after having analysed both the main and subcategories. Although having been developed in an expert-novice dialogue with direct communication and feedback, the application on online explanatory videos from the video-sharing site YouTube by the measure can thus be seen as an additional application possibility since the measure seems to be able, although not proven significantly due to the small sample size, that it is basically able to predict learning outcomes. This result is alignable with Wolf's findings that already listed similarities in the preparation between explanatory videos and a situation when explaining something to somebody face-to-face. These similarities may include addressee-adequacy by using an appropriate level of speech, considering the addressee's prior knowledge as well as attitudes and interests. Producing an explanatory video can be considered the more demanding task as the explainer does not receive instantaneous feedback. Even though the subcategory *addressing pre- and misconceptions* appears eight times directly (Figure 15), it is still part of a lot of other subcategories such as using *examples close to everyday life*, different *contexts*, and *paraphrasing technical terms in everyday language*. Moreover, the measure is also able to mark differences at videos produced for use at professional teaching environments (U-videos) and videos that are produced for private disseminating purposes (P-videos), with U-videos accumulating slightly more CPs on average (10CPs) than P- or T-videos, that accumulate six and eight CPs on average respectively (Figure 12). P-videos, though, offer among summaries that use formulae an approach that appears both more graphic and descriptive than U-videos by using outlooks, examples close to everyday life and non-verbal media such as logical figures and writing (Figure 14 and 15). U-videos, on the other hand, offer clear structure through summaries, emphasising important points and justifying approaches (Figure 14 and 15) as well as their explainers generally use more technical terms which are usually commented close to the subject or close to everyday life as seen in Figure 15.

The question whether the measure, which is based on communication models with two active parties involved, can be used on settings without immediate feedback and one active party only can therefore be affirmed. All four variables from the communication model (Mathematization, Code, Graphic Presentations, Examples and Analogies) reappear as well as seven out of ten quality criteria also remain incorporated into the 31 used subcategories (see Table 4). That means that the scaffold of the measure based on the 45-category catalogue remains sufficiently intact to begin further investigations. Furthermore, relations to Merten's and Rusch's communication models can also be derived still as viewers must actively select (Merten, 1995) or can ignore (Rusch, 1999) videos or their content. Hence, the viewer maintains the basic active role he or she plays in their respective communication models: Viewers listen, select and change videos as necessary to find better or more appropriate explanations suitable to their individual learning needs. They may look for videos with, for instance, more (or less) examples closer to their experience, more illustrative figures, less or more mathematization to name a few. Brown (1986) emphasised the need of an explanation to be part of the addressee's environment, that is why examples ought to be close to everyday life. The idea of constructivism and addressee-adequacy as well as considering prior knowledge seem to be present at almost all science teaching explanations, consequently they seem to be omnipotent characteristics of science explanations. This may also make them indispensable when amending communication models to investigate explaining skills. A possible drawback from giving the explainee so much control over choosing an explanation that suits his or her personal needs might result in picking videos that may fit neatly into the explainee's inner constructed world but may feed on flaws and wrongs, thus leading the viewer onto a path that may become hard to leave as it solidifies his or her false perceptions and constructs of the world.

Choosing Kelper's Laws as the videos' topic proved to be the right choice since it is suitable for explanations that do not require a whole lot of mathematics and forces the explainer to use a wide range of examples, contexts, and numerical examples. The methods both frequency quantitative and inductive qualitative content analysis as well as statistics to compare pre- and post-test means were specific and target-oriented to reach the findings efficiently and objectively. The choosing of some YouTube surface features appear arbitrary, as **Likes** and **Views** were included because YouTube simply provides those data. Other data, on the other hand, like **Average View Duration** and **Relevant Comments** were chosen a lot more deliberately. More care should therefore have been used at selecting concurrent data: For example, including **Shares** or the number of **Channels** or **Followers** of a YouTube publisher might have given better or at least different insights. Furthermore, the test and questionnaire, although resulting in feasible results, could have been more extensive and varied if more time had been at disposal. Equally, using a video for learning might be beneficial for some but not at the least helpful to others, thus using an entire group without individualizing the process of learning might have been felt obstructive to some participants. To improve data collection of the field study, a lot more time is needed to grant the students individual time and manipulation of the video and allow them to more actively control their personal learning process. In the same manner, rather than simply asking students to recite facts that are part of an explanatory video, the application of newly gained knowledge on contexts and examples might be a more suitable way to determine learning impact, and whether it shows they have actually constructed and incorporated new facts successfully into their existing construct of the world. This claim is supported by Brown who recommends to *"apply [an] explanation to another situation or related problem"* (Brown, 1986, p. 203) to gauge comprehension.

As a result, it seems more likely that educators, teachers, and people who look for an explanation that offers a good explaining quality might find it (a) by reviewing the number of content-related *Comments* and/or (b) by watching the video and applying the measure's 45 subcategories or, alternatively, the 12 core categories. It also seems viable to choose between a more structures (U-video) or a more descriptive video (P-video).

This likelihood seems to be a fair starting point to begin new investigations into the measure's qualities by having, for example, layman or experts use **Relevant Comments**, and the twelve core categories to find rich and varied explanations or to distinguish between two science explanation videos concerning their explaining quality and thus the explainer's explaining skill.

5.1 Outlook and Improvements

This work has peered into a relatively new direction of quantifying explaining skills and has probably uncovered more questions that it is able to answer satisfactorily. As Brown has already noted since 1986, see (2.1), students need feedback to re-evaluate their constructed ideas and thus, it might be interesting to observe students' behaviour while they are searching and selecting (science) explanatory videos, and whether they use different videos against each other to validate their respective content for correctness rather than to be inclined in giving a comment. This behaviour study will be of importance as the viewer is still in an active role during this process but he or she does not have to, or better, cannot give direct feedback by asking questions during the explanation or being involved by being asked questions by the explainer. It might therefore be interesting to videotape viewers' activities to evaluate for how long they interact with explanatory videos in a more natural setting rather than showing a whole group the same video once. Selectivity, above all, is an important process while learning mentioned by Merten (see Chapter 2.1). Also, choosing from a huge pool of explanatory videos is both YouTube's biggest advantage (Wolf, in press) and challenge because viewers have to find an efficient and, at the same time, easy way to navigate **and** to choose a suitable video from YouTube's large database. Apart from YouTube's *Up-Next* feature, which must be considered highly manipulative, the users are left alone in their decision-making process. This is why the question by what students are influenced in their choice is important to answer. In addition, Wellenreuther (2013) mentions students' possible trouble to have their working memory overloaded, especially during complex explanations. Explanatory videos could be seen as an additional tool to slow down the learning process while allowing the learner to adapt his or her individual learning pace. This is also the main advantage mentioned for blended-learning environments. Exploring this special question might shed some light on their application in the modern age as most educators are still uncertain about how to approach them efficiently, apart from putting the student into the centre of all considerations (Mandl & Kopp, 2006).

Similarly, to gauge the quality of the measure even further, it must be exposed to new topics and videos to prove its general suitability, which was determined for science explanatory videos on Kepler's Laws exclusively, thus improving its criteria validity. Also, rather than having students' learning impact decide on its predictive validity, experts' decisions in ranking superior versus inferior videos ought to be considered to solidify the application of the measure and a more arbitrary set of explanations than sole expert-novice dialogue situations.

Moreover, the measure was successful in ranking science explanations quality, but maybe these 12 core categories, or all 45 for this matter, can also be used to *produce* more powerful and convincing explanations? The production of science explanations videos can be used, for example, to mark the introduction of a new topic, be used to continually document someone's learning success, or as a recap (Wolf, in press), and whether self-producing explanatory videos may affirmatively affect students' attitude towards the subject, as discovered by Chandra and Watters (2012).

Of no less interest might be the reason for some people's motivation to produce such an explanatory video and to publish it at no charge. Further studies are also required to examine the structure of explanations to answer questions whether it may be possible to find certain patterns that reoccur in explanations with a high or low CP rank.

This short selection of newly emerged questions illustrates the need for further research on explanatory videos which cannot be overstressed as even schools, official school administration bodies and authorities offer links to explanatory videos on their servers, as found on Lower Saxony's website which provide teachers with links to *LEIFI* and *Siemens* (http://www.nibis.de/nibis.php?menid=6677). DIE ZEIT has recently urged educators to realize that libraries and lectures may disappear altogether as few clicks on YouTube already show ubiquitous educators (ZEIT, 2015b). DIE ZEIT also dares to claim that *"the entire biological, cognitive, medial, academical apparatus of knowledge production and distribution stands before its biggest turnover since Humboldt's university reforms and the discovery of printing"* (ZEIT, 2015b, p. 69, translation by the author). Mr. Salman Khan, founder of *The Khan Academy*, also believes *"that the way how we teach and learn has reached an inflection point, that kind we witness only once within a thousand years"* (Khan, 2013, p. 9, translation by the author). Remembering Merten's quote from page one makes me fear that this turnover has already gained pace while having been fairly overlooked by the teaching community.

This work is supposed to be considered a first step to establish some order in the quality of some of the ubiquitous information that, as TIME mentioned, is only some keystrokes away.

6 Bibliography

- Alexander, S. (2010). Flexible Learning in Higher Education. In P. Peterson, E. Baker & B. McGaw (Ed.). *International Encyclopaedia of Education (Third Edition)*. Oxford: Elsevier, pp. 441-447.

- Altman, D.G. (1997). *Practical Statistics for Medical Research.* Boca Raton: Chapman & Hall.

- Barcelona Field Studies Centre (2015). *Significance of Spearman's Rank Correlation Coefficient.* Available at: http://geographyfieldwork.com/SpearmansRankSignificance.htm, 22/June/2015

- Bartelborth, T. (2007). *Erklären.* Berlin: De Gruyter.

- Bergmann, M. (Ed.) (2001). *Duden Physik. Ein Lexikon zum Physikunterricht.* Mannheim: Dudenverlag.

- Brown, G. (1986). Explaining. In: O. Hargie (Ed.). *A Handbook of Communication Skill.* Worcester: Billing & Son Ltd.

- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion.* München: Pearson.

- Chandra, V. & Watters, J.J. (2012). Re-thinking physics teaching with web-based learning. *Computers and Education, 58*, (1), pp. 631-640.

- Clark, I. (2005). *Blended learning: an approach to delivering science courses on-line.* Symposium Presentation. University of South Australia-Mawson Lakes: UniServe.

- Davis, O. (2015). *Inter-Rater Reliability in Psychology: Definition, Formula &*

Quiz. Available at: http://study.com/academy/lesson/inter-rater-reliability-in-psychology-definition-formula-quiz.html, 13/June/2015

- Diekmann, F., Gleiche, O., Weiner, A. (2010). Interesse wecken an ingenieurwissenschaftlicher Tätigkeit. Unterstützung technologischer Experimente durch internet-gestützte Medien. In G. Kammasch, A. Schwenk, B. Wieneke-Toutaoui (Ed.). *IngenieurBildung für Nachhaltige Entwicklung* (pp. 66-71). Berlin: Beuth Hochschule für Technik.

- Duit, R. (2004). *Schülervortsellungen und Lernen von Physik*. Piko-Brief Nr. 1. Kiel: IPN.

- Fleiss, J. L., Levin, B. & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions* (3rd ed. Aufl.). Wiley Series in Probability and Statistics. Hoboken, New Jersey: J. Wiley.

- Garisson, D. R. & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential i n higher education. In *The Internet and Higher Education*, (7), pp. 95-105.

- Google Developers (2015). *YouTube Analytics API*. Available at: https://developers.google.com/youtube/analytics/v1/dimsmets/mets, 11/June/2015

- Grouven, U., Bender, R., Ziegler, A. & Lange, S. (2007). Der kappa-Koeffizient. In *DMW - Deutsche Medizinische Wochenschrift* 132 (S 01), pp. e65-e68.

- Hammann, M., Jörgens, J., Schecker, H. (2014). Übereinstimmung zwischen Beurteilern; Cohens Kappa (κ). In D. Krüger, I. Parchmann & H. Schecker (Ed.), *Mathoden der naturwissenschaftlichen Forschung*. Berlin: Springer.

- Khan, S. (2013). *Die Khan Academy*. München: Riemann.

- Kleinhanß, C. (2015). Erklärvideos. In *Computer + Unterricht*, *Vol. 25*, (97), p. 41.

- Kulgemeyer, C. & Schecker, H. (2013). Students Explaining Science - Assessment of Science Communication Competence. In *Research in Science Education 43*, pp. 2235-2256.

- Kulgemeyer, C. & Tomczyszyn, E. (in press). *Physik erklären können - Diagnostik adressatengemäßer Erklärkompetenz angehender Physiklehrkräfte.*

- Langer, W. (1999). *Methoden V: Explorative Faktorenanalyse.* Available at: http://www.soziologie.uni-halle.de/langer/lisrel/skripten/reliabil.pdf, 15/June/2015

- Lani, J. et al (2015). *Kendall's Tau and Spearman's Rank Correlation Coefficient.* Available at: http://www.statisticssolutions.com/kendalls-tau-and-spearmans-rank-correlation-coefficient/, 08/July/2015

- LaPointe, Evan (2010). *How does YouTube video view count work?* Available at: http://www.atlantaanalytics.com/practicing-web-analytics/how-does-youtube-video-view-count-work/, 10/June/2015

- Mandl, H., Kopp, B. (2006). *Blended Learning: Forschungsfragen und Perspektiven.* (Forschungsbericht Nr. 182). München: Ludwig-Maximilian-Universität, Department Psychologie, Institut für Pädagogische Psychologie.

- Mayring, P. (2000). Qualitative Content Anaysis. In *Forum: Qualitative Social Research.* Vo. 1, No.2.

- Merten, K. (1995). Konstruktivismus als Theorie für die Kommunikationswisenschaft. In: *Medien Journal*, 4, pp. 3-21.

- Owen, A., Matthews, A.,Petrie, M., Green, D., Croft, T. (2015). *Spearman's correlation.* Available at: http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf, 22/June/2015

- Parsons, J. (2014). *How Does YouTube Calculate Your Video Views.* Available at: http://boostlikes.com/blog/2014/05/youtube-calculate-video-views, 08/June/2015

- Rusch, B., Friese, M., Hofmann, W., Naumann, E. (2006). *Quantitative Methoden.* Band I. Berlin: Springer.

- Renkl, A. (1996). Lernen durch Erklären - oder besser doch durch Zuhören? In *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, XXVIII* (2), pp. 148-168.

- Reinmann-Rothmeier, G., Mandl, H. (2001). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidenmann (Ed.), *Pädagogische Psychologie* (pp. 601-646). Weinheim: Beltz.
- Reinmann-Rothmeier (2003). *Didaktische Innovation durch Blended-Learning. Leitlinien anhand eines Beispiels aus der Hochschule.* Bern: Huber.
- Rusch, G. (1999). Eine Kommunikationstheorie für kognitive Systeme. Bausteine einer konstruktivistischen Kommunikations- und Medienwissenschaft. In: *Konstruktivismus in der Medien- und Kommunikationswissenschaft.* Frankfurt am Main: Suhrkamp, pp. 150-184.
- Schecker, H. (2007). *Erklären im Physikunterricht.* Unpublished presentation, University of Bremen.
- Schecker, H., Parchmann, I. & Krüger, D. (2014). *Methoden in der naturwissenschaftsdidaktischen Forschung.* Berlin: Springer.
- Schüßler, I. (2004). *Lernwirkungen neuer Lernformen.* Berlin: Arbeitsgemeinschaft Betriebliche Weiterbildungsforschung e. V.
- Senator für Bildung und Wissenschaft (Ed.).(2006). *Naturwissenschaften, Biologie - Chemie - Physik. Bildungsplan für das Gymnasium.* Jahrgangsstufe 5-10. Bremen.
- Stangl, W. (2015). *Empirische Forschungsmethoden in Pädagogik und Psychologie.* Available at: http://arbeitsblaetter.stangl-taller.at/Forschungsmethoden, 20/June/2015
- The Guardian (2014). *Behind the online comments: the psychology of internet trolls.* Available at: http://www.theguardian.com/media-network/media-network-blog/2014/sep/18/psychology-internet-trolls-pewdiepie-youtube-mary-beard, 08/June/2015
- TIME (2006). *YOU. Person of the Year.* Vol. 168, NO. 26, cover.
- TIME (2015). *What's this all about.* Vol. 186, NO. 1-2, pp. 36-37.
- Tomczyszyn, E., Kulgemeyer, C. & Schecker, H. (2014). Diagnostik des Erklärungswissens von Physikstudierenden. In Bernholt, S. (Ed.), *Naturwissenschaftliche*

Bildung zwischen Science- und Fachunterricht. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in München 2013 (pp. 114-116). Kiel: IPN.

- University of Dallas (2013). *Helping Students Learn Physics Better.* Available at: http://phys.udallas.edu/C3P/Preconceptions.pdf, 11/June/2015

- Van (2013). *What counts as a view on YouTube?* Available at: http://www.slideshare.net/van_ slides/what-counts-as-a-view-on-youtube, 11/Juni/2015

- von Thun, F. (1998). *Miteinander Reden - 3 - Das "Innere Team" und situationsgerechte Kommunikation.* Reinbek: Rowohlt.

- Welbourne, D.J.; Grant, W.J. (2015). *Science communication on YouTube: Factors that affect channel and video popularity.* Puplic Understanding of Science, pp. 1-14. Australian National University.

- Wellenreuther, M. (2000). *Quantitative Forschungsmethoden in der Erziehungswissenschaft.* Weinheim und München: Juventa Verlag, pp. 270-283.

- Wellenreuther, M. (2013). *Lehren und Lernen - aber wie?* Hohengehren: Schneider Verlag.

- Wolf, K.D. (in press). *Bildungspotenziale von Erklärvideos und Tutorials auf YouTube: Audio-Visuelle Enzyklopädie, adressatengerechtes Bildungsfernsehen, Lehr-Lern-Strategie oder partizipative Peer Education?* Universität Bremen.

- Wolf, K.D. (2015). Video-Tutorials und Erklärvideos als Gegenstand, Methode und Ziel der Medien- und Filmbildung. In Trültzsch-Wijnen, C. & Hartung, A. (Ed.), *Filmbildung im Wandel.*

- Wolf, K.D. & Kratzer, V. (in press). Erklärstrukturen in selbsterstellten Erklärvideos von Kindern. To be published in *Jahrbuch Medienpädagogik.*

- ZEIT (2015a). *Halbwissen - Quelle: Internet.* № 28, 9/July/2015

- ZEIT (2015b). *Lernt von den Studenten!*, № 25, 18/June/2015

7 List of Videos and Related Data

This chapter shows all collected raw data used for further calculations. All data were collected at the beginning of May 2015. For its exact collection method, see 3.3.

Table 7: This table (1/2) shows all data used for calculations. All data as of May 2^{nd}, 2015.

Video Code	Views	CP	Months online	Run-Time	Average View Duration	Total Likes	Total Dislikes	Total Comments	Total Relevant Comments	URL
DP05	6737	6	26	246 s	113 s	37	8	10	3	https://www.youtube.com/watch?v=j0hsERNkz1U
DP11	1230	8	21	657 s	140 s	8	0	1	0	https://www.youtube.com/watch?v=tWCosFFf3cY
DT04	386	6	9	217 s	128 s	1	0	2	0	https://www.youtube.com/watch?v=g0MRPJgZaiE
DT15	17657	5	50	869 s	-	40	3	12	2	https://www.youtube.com/watch?v=jMgnV8_ eew8
DU16	16574	10	31	954 s	-	198	1	25	7	https://www.youtube.com/watch?v=pI5SfiD0WeU
DU20	12056	11	28	1194 s	146 s	28	8	1	1	https://www.youtube.com/watch?v=rhD110v5nyY
EP02	190	3	4	95 s	52 s	0	0	0	0	https://www.youtube.com/watch?v=oE5sUAdxlxM
EP02dd	613	6	4	108 s	-	2	0	0	0	https://www.youtube.com/watch?v=qUTa8reldYI
EP03d	518	3	4	154 s	-	1	0	0	0	www.youtube.com/channel/UCEyPGpoUc9-q-G2N4sNwiaQ
EP04	296	9	19	201 s	-	1	0	2	1	https://www.youtube.com/watch?v=T1yEeFbfhkY
EP04b	10855	10	14	204 s	131 s	58	1	11	6	https://www.youtube.com/watch?v=92ROKc-BwDc
EP04c	53402	7	64	189 s	-	183	21	52	5	https://www.youtube.com/watch?v=WTJa8DXlerc
EP04dddd	593	8	4	228 s	-	5	0	0	0	www.youtube.com/channel/UCEyPGpoUc9-q-G2N4sNwiaQ
EP04g	636	13	5	224 s	140 s	7	3	9	9	https://www.youtube.com/watch?v=s77LJO6USEY
EP05ddd	636	11	4	296 s	-	5	0	0	0	https://www.youtube.com/watch?v=5TyENmdHzlk
EP06	16471	8	46	318 s	-	69	5	17	7	https://www.youtube.com/watch?v=XBxDDXG3sTQ
EP07	6304	2	39	370 s	-	11	6	6	3	https://www.youtube.com/watch?v=6rv9oMXDPMw
EP13	6522	11	30	728 s	388 s	62	1	35	10	https://www.youtube.com/watch?v=zNeFI_ JCXlY
EP13f	101574	15	40	753 s	-	806	16	178	62	https://www.youtube.com/watch?v=6TGCPXhMLtU
EP16	352	8	19	901 s	251 s	3	0	2	1	https://www.youtube.com/watch?v=blRc6NuFh-Y

Table 8: This table (2/2) shows all data used for calculations. All data as of May 2^{nd}, 2015.

Video Code	Views	CP	Months online	Run-Time	Average View Duration	Total Likes	Total Dislikes	Total Comments	Total Relevant Comments	URL
ET04	324	6	36	201 s	-	7	2	0	0	www.youtube.com/channel/UCvQhV0impINSA08m-Dv_ l-g
ET05	4790	6	69	249 s	-	17	0	14	3	https://www.youtube.com/watch?v=XFqM0lreJYw
ET06	82850	6	77	352 s	-	247	7	69	9	https://www.youtube.com/watch?v=GcKiG-CuvtA
ET15	16048	3	13	868 s	390 s	64	2	15	4	https://www.youtube.com/watch?v=4p3Np6eivZ8
EU02	68	3	7	83 s	57 s	0	0	0	0	www.youtube.com/channel/UCzoUwL8Yw0qDp-CC0oNdy3w
EU04e	12279	12	33	191 s	-	89	4	40	4	https://www.youtube.com/watch?v=CKsf9XpzclY
EU05	8382	5	24	269 s	102 s	11	3	2	0	https://www.youtube.com/watch?v=4wrTHOCRxy4
EU05a	3783	8	40	260 s	-	21	1	4	1	https://www.youtube.com/watch?v=Z4qgKoVDC4c
EU06ee	10781	9	33	338 s	-	113	1	22	5	https://www.youtube.com/watch?v=v6yjGYIJRFw
EU07	10984	10	76	414 s	-	50	1	10	2	https://www.youtube.com/watch?v=wjOOrr2uPuU
EU07aaa	14195	11	40	413 s	-	62	4	5	4	https://www.youtube.com/watch?v=FjAdqr1Qbac
EU09	3558	4	39	534 s	-	13	1	4	0	www.youtube.com/channel/UCH_ PiMqviAn_ nlGez8_ PmjQ
EU09aa	11295	7	47	512 s	-	68	3	20	10	https://www.youtube.com/watch?v=Pa3Of_ 3vpRc
EU09k	14178	8	63	519 s	-	52	5	20	4	https://www.youtube.com/watch?v=QsopVzjd2sU
EU10	4638	10	49	600 s	-	4	2	5	2	www.youtube.com/channel/UClShfV0zAprEYxtVYQT7WDg
EU10h	179	12	9	544 s	308 s	1	0	2	2	https://www.youtube.com/watch?v=7ZQym42wJFU
EU12	291	9	14	669 s	-	0	0	1	0	https://www.youtube.com/watch?v=vsRe6W-cWms
EU12i	1413	7	9	714 s	122 s	2	1	0	0	https://www.youtube.com/watch?v=WA1H98-bEfE
EU22	119	15	13	1297 s	425 s	1	0	3	0	https://www.youtube.com/watch?v=yeJgfqZkmio
EU27	1946	18	29	1615 s	371 s	4	0	2	2	https://www.youtube.com/watch?v=Kyvhonwu9Ss

8 List of Figures

List of Figures

8.1 Knowledge Test

- Question 3: *Lissajou Figure.* Available at: http://jan.ucc.nau.edu/ jws8/ 137.2001.1/lissajou3d2still.gif [11/June/2015]
- Question 3: *Hysteresis Loop.* Available at: https://lp.uni-goettingen.de /get/text/3792 [11/June/2015]
- Question 3: *Ellipse.* Available at: http://catalog.flatworldknowledge.com/ bookhub/reader/4372?e̅fwk-redden-ch08 _ s03 [11/June/2015]
- Question 3: *Closed Curve.* Available at: http://texwelt.de/wissen/fragen/3379/wie-zeichnet-man-beliebige-zusammenhangende-gebiete-bzw-beliebige-funktionen [11/June/2015]

9 List of Tables

List of Tables

10 Annex

Table 4 Characteristics of categories of explaining identified in the study

	Category	Characteristics	Example in scenario "liquid expansion thermometer"
Cognitive categories	Give examples	Ability to use appropriate examples and change them (if required) into everyday related examples	"This is the same with bridges for example, they need room to expand."
	Use graphs	Ability to use given graphical representations to visualize the explanation	
	Produce graphs	Ability to apply graphical representation forms to visualize the explanation	
	Link graphs	Ability to link multiple graphical representation forms with each other to visualize different aspects	"Look at this picture… and now look at this sheet…"
	Vary the model	Ability to vary the scientific model used in the explanation	Switch between a particles model and a continuum model
	Vary the level of abstraction	Ability to vary between different levels of abstraction	Switch between anthropomorphic descriptions (e.g. particles as humans) and scientific models
	Vary code	Ability to change the used code from a scientific level into everyday language if the addressee asks for it	Change from "the coefficient of linear expansion is…" to "the expansion happens just like …"
Content knowledge	Give concise answers	Direct answers to questions about scientific terms or definitions	"What are particles?"—"They are these tiny little thingies a fluid consists of."
Volitional categories	No interruptions	No interruptions while addressee is talking	
	Confirm understanding	Asking the addressee whether she/ he has understood what has been explained	"Did you get that?"
	Direct addressing	Ability to use commented scientific terms and address the younger student directly	Using the second person singular ("you") or the imperative
	Inquire prior knowledge	Ask the addressee what he/she already knows about a phenomenon or the factual background	"What do you know about particles?"
	Inquire needs	Finding out what the addressee wants to have explained	"Anything else you'd like to know?"
	Prepare an introduction	Evidence for a prepared first explanation for the start of the role play; point out the relevance of the phenomenon	"The fluid thermometer is a commonly used instrument to measure the temperature, you've probably got one at home yourself. It works as the fluid in it expands…"
	Inquire prior knowledge initially	Inquiring the addressee's prior knowledge at the beginning of the role play	"I want to explain the way a thermometer works. What do you already know about it?"
	Introduce the topic	Naming the topic at the beginning of the role play	"Hi, my name is Rieke, and I'd like to explain to you how the thermometer works."

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
	Proband → Inhalt	
Inhalt - falsches Thema	An den Äußerungen des Erklärers wird deutlich, dass dieser die Situation im Szenario nicht verstanden hat. Es wird nicht das Erklärt, was erklärt werden soll 1 Der Proband hat das Phänomen/die Situation falsch Verstanden. 2 Der Proband erklärt eine physikalische Größe, mit der man das Szenario nicht erklären kann - unabhängig davon, ob diese richtig dargestellt wird. Die Größe hat nichts mit dem Szenario zu tun.	1 z. B. man fühlt sich bei der Achterbahn nur am höchsten Punkt schwerelos. • Die Form der Hügel ist egal, um sich schwerelos zu fühlen. 1 das U-Boot ist nur deswegen so schwer zu bergen, weil so viel Wasser drin ist ??? 2 Der Impuls wird verwendet, um die Knautschzone zu erklären (Pilot_1_01a 1:49) • Redet z. B. über ein anderes Thema, damit Zeit verstreicht.

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Inhalt - Fehler	Prinzipiell wird über das Richtige gesprochen aber der Proband macht Fehler. • Beim Erklären des richtigen Konzepts (d. h. keine Markierung bei „Inhalt - falsche Erklärung“) werden Fehler gemacht. Das heißt es wird über die richtige physikalische Größe gesprochen aber es werden Fehler gemacht. • Ein physikalisches Konzept oder ein Begriff wird falsch erklärt oder eine falsche Definition genannt. • Die Fachbegriffe werden falsch verwendet (Nicht wenn andere Begriffe falsch genannt werden: Z. B. Autobahn statt Achterbahn Prä_1_09) • Ein Alltagsbegriff wird mit einer physikalischen Bedeutung belegt und dann falsch verwendet. • Eine Abbildung wird falsch interpretiert. • Beim Sprechen über die Zentrifugalkraft wird der Wechsel des Bezugssystems nicht deutlich gemacht. Wird hier eingeordnet, weil es möglich ist die Kurvenfahrt mit der Zentrifugalkraft richtig zu erklären.	• Gegenstück mit der Bedeutung Gegenkraft (Prä_1_03 5:50) (6:10)„Das ist das Gegenstück zur Zentrifugalkraft, die Zentripetalkraft“ (6:55) „F_Z ist die Zentripetalkraft. Das ist das Gegenstück zu der Zentrifugalkraft.“ • Abbildung falsch bei Pilot_1_01a ab 3:46, 9:21 und 9:54. • (Prä_1_03 5:03) „Das was wir spüren ist diese Scheinkraft, die Zentripetalkraft“ • „Wenn diese Reibungskraft nicht groß genug ist, wird das Motorrad durch die entgegengesetzte Kraft nach außen geschleudert.“ (Prä_1_03 3:20) • Wechselwirkungsprinzip wird mit einem Spiegel verglichen (Pilot_1_01a 3:17)
Inhalt - Fehler korrigieren	• Ein Fehler wird durch den Probanden explizit korrigiert. Er/Sie bemerkt den Fehler selbst oder wird vom Adressaten darauf Aufmerksam gemacht. • In diesem oder einem vorherigen Segment wurde die Kategorie „Fehler“ markiert. • NICHT: Offensichtlicher Versprecher, der sofort korrigiert wird.	P: „...entsprechend der Masse könnte man genau berechnen, wie groß der Impuls für das Obere ist und für das Untere.“ A: „Aber ich dachte der muss gleich sein.“ P: „P: Ja, man weiß dann wie groß die Geschwindigkeit ist, ne.“
	Proband → Medium → Verbal → Strukturierung der Erklärung Teil 1	

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Ausblick (Advance Organizer)	Der Proband gibt explizit, kurz und zusammenhängend einen Ausblick auf: den physikalischen Inhalt, den Ablauf oder das Szenario/Thema. 1: Der Proband sagt explizit, welchen physikalischen Inhalt er erklären will. 2: Sagt explizit etwas über den Ablauf der Erklärung. 3: Stellt das Szenario/Thema explizit, richtig, vollständig und auf den Punkt gebracht vor. (Achterbahn: Form der Schienen. Asteroid: Flugbahn nach der Explosion. Kurvenfahrt: Unfall bei nasser Kurve. U-Boot: Schwierigkeit es zu bergen. Knautschzone: Fahrgastsicherheit durch Knautschzone.)	1: „Ich versuch dir jetzt zu erklären woher das [Schwerelosigkeitsgefühl] kommt.“ Prä_1_08 0:44 (Ähnlich: Pilot_1_03a 0:11) 1: „...habe ich mir zunächst überlegt, dass ich dir erkläre: Was ist Geschwindigkeit und was ist Beschleunigung.“ Pilot_1_11a 0:36 2: „Ich komme gleich darauf zurück. ich mache erst mal ein bisschen weiter...“ Pilot_1_04b 2:42 2: Ich erkläre dir zuerst... und dann... 3: „Und ich erkläre dir jetzt, warum das [Unfälle in nasser Kurve] so oft passiert ist.“ Prä_1_03 0:30 3: „... Aber das wollen wir ja gerade ermitteln, wie das [Bahn des Asteroiden] zu Stande kommt.“ Prä_1_02 3:35 • NICHT: Jetzt geht es um Asteroiden.
Rückblick (Strukturierend)	• Der Proband sagt explizit, dass er etwas vorher schon mal zeigt oder erklärt hat. Auf Dinge, die vorher gesagt oder gezeigt wurden, wird mit einer Äußerung zurückgegriffen. Aber der Inhalt wird nicht nochmal erklärt.	• „Du hast ja gerade von der Zentrifugalkraft gesprochen...“ Prä_1_03 5:35 • „In der Kurve passiert genau das, was ich vorhin mit der Kreisbewegung gemeint habe...“ Pilot_1_01a 5:36

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Zusammenfassung / Wiederholung	• Zusammenfassung oder Wiederholung eines Sachverhalts, der schon erklärt wurde. • Es wird explizit herausgestellt, dass es sich um eine Zusammenfassung oder Wiederholung handelt, indem der Proband sagt, dass der Inhalt schon besprochen wurde.	• „Wir haben ja am Anfang besprochen, dass man auf nasser Straße einen längeren Bremsweg hat und auf vereister Straße einen noch längeren Bremsweg. Und auf nasser Straße hat man einen längeren Bremsweg, weil die Reibung nicht so groß ist...“ Prä_1_03 8:00
Ermunterung	• Der Proband lobt den Adressaten, ermuntert ihn etwas zu sagen oder zu tun, oder äußert Verständnis für die Schwierigkeiten, die der Adressat hat. • Der Adressat wird dabei direkt angesprochen.	• „Ich weiß, klingt schwer am Anfang...“ (- das „am Anfang ‘“ deutet auf den Adressaten hin) Pilot_1_01a 5:36 • „Wenn du irgendwas nicht verstehst, musst du nachfragen“ Pilot_1_03a 7:00 • „Ich weiß, das ist viel...“ Pilot_1_01a 4:32 • (NICHT: Pilot_1_05a 4:40 und 8:00 Da wird nur gesagt, dass das Thema schwer ist)
Herablassende Äußerungen	• Der Proband macht eine herablassende Äußerung über das Wissen oder die Fragen der Adressaten. • Der Tonfall ist unangemessen.	• „Aber ich glaub ich mach jetzt heute keine Berechnung“ Pilot_1_06b 8:06
Bedeutungshinweise geben	• Der Proband sagt explizit, dass ein Sachverhalt wichtig ist, gemerkt oder verstanden werden muss.	• „Das ist erst mal wichtig zu Wissen...“ Pilot_1_11a 0:42 • „Merk dir einfach...“ Pilot_1_11a 2:14 • „Was dabei klar sein muss...“ Prä_1_02 3:40

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
begründet Vorgehen	• Der Proband begründet sein Vorgehen sinnvoll. • Lehnt ab auf einen Impuls des Adressaten einzugehen, weil implizit vorausgesetzt wird, dass Wissen oder Verständnis fehlt.	• Pilot_1_06a 8:10 Mir geht es erst mal darum, dass man so ein Gefühl dafür kriegt... Es ist eigentlich schöner, wenn man sich erst mal was vorstellen kann. • „Mein Ziel war es dir erst mal allgemein zu erklären...“ Prä_2_16 8:47
direktes Ansprechen von Schülervorstellungen	• Proband macht deutlich, dass er eine typische Schülervorstellung erkannt hat oder dass er eine Schülervorstellungen kennt. • Z. B. indem es explizit eingebaut wird: „Viele Schüler denken...“	• Pilot_1_08b 5:43 „Manche würden vielleicht sagen... aber das stimmt nicht...“ • Du denkst wahrscheinlich ..., aber ... • In der Physik hat das eine andere Bedeutung als im Alltag.
	Proband → Medium → Verbal → Vortrag *Vortrag sind Aussagen, die mehr als nur Zustimmung (Ja, Stimmt, Richtig...) oder Ablehnung (Nein, Nicht ganz...) einer Adressatenäußerung sind. Es muss mindestens der Inhalt genannt werden, dem Zugestimmt/ der Abgelehnt wird.* *Diese Kategorien bei Fragen NICHT kodieren.*	
Antropomorphisierungen	• Gegenständen werden menschliche Eigenschaften zugesprochen. • Unbelebte Gegenstände haben einen Willen.	• „Die Geschwindigkeit ... die das System selber aufnehmen möchte.“ Pilot_1_11a 2:58

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
(nur physikalische Fachbegriffe)	Immer als Fachbegriff kodieren: • Geschwindigkeit, Geschwindigkeit mit Richtungsangabe (auch v_x, v_y etc.), gleichförmige Bewegung • Beschleunigung, gleichmäßig beschleunigte Bewegung, Kreisbewegung • Kraft • Impuls • Actio und Reactio, Wechselwirkungsprinzip • Arbeit, Energie • freier Fall (fällt frei etc.), schräger Wurf (waagerechter Wurf etc.) • Schwerelosigkeit • System, Masse, Massenpunkt • Reibung, Haftreibungskoeffizient • Trägheit • Druck, hydrostatischer Druck • Einheiten (Newton oder N oder zusammengesetzte Einheiten km/h etc.) Vielleicht doch nicht als Fachbegriff kodieren, weil das Alltagskonzept dem Physikalischen entspricht: • Objekt, Ort, Strecke, Weg, Zeit, Gewicht (und die Einheiten davon)	*Kodiert werden:* • *Fachbegriffe* • *Kompositionen mit einem Fachbegriff* • *Abkürzungen und Formelzeichen, die für Fachbegriffe stehen.* • *Pronomen, die für einen Fachbegriff stehen „Die [gemeint ist Geschwindigkeit] wird immer schneller (Prä_08 5:41-6:07)*

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Umschreibung von Fachbegriffen	• Der Proband vermeidet Fachbegriffe, indem er diese durch Alltagsbegriffe ersetzt. • Eine Kommentierung eines Fachbegriffs wird genannt, ohne den Fachbegriff an sich zu nennen.	• Geschwindigkeit: „schnell sein„, „sich in die Richtung bewegen“ • Beschleunigung: „schneller bzw. langsamer werden“, „bremsen“ • Impuls: „Schwung“ • Reibung: ”Haftung, Halt ‘“ • Masse: Gewicht, schwer, schwerer, leicht, leichter, einfacher • Pilot_1_05a 1:07 Erdanziehungskraft: „Wenn du hier ein Gewicht hängen hast, dann wird es ja von der Erde angezogen.“ • Pilot_1_05a 1:58 Auftrieb: „... und gleichzeitig wird es vom Wasser bisschen angehoben“
kommentierte Fachbegriffe (alltagsnah)	• Kommentierung, indem Eigenschaften, die den Fachbegriff definieren und von anderen abgrenzen, direkt mit Alltagssprache beschrieben werden (Inhaltsbeschreibung). Ohne davor eine Situation zu beschrieben. • Ein Alltagsbegriff mit ähnlicher Bedeutung wird genannt und als Äquivalent beschrieben (Ausdrucksersetzung) NICHT: Begriffe Synonym verwenden, ohne zu sagen, dass sie es sind. • *Allein durch die Kommentierung muss eindeutig sein, welcher Fachbegriff gemeint ist.*	• Unter Impuls kannst du dir so etwas wie Schwung vorstellen. • Pilot_1_01a „Die Masse deines Körpers, dein Körpergewicht...„ • „Eine Kraft bewirkt, dass du deine Bewegung änderst“ Pilot_3_01b 3:50 • (Siehe: Kommentierung von FB.docx)

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
kommentierte Fachbegriffe (mit Fachbegriffen)	• Kommentierung, indem Eigenschaften, die den Fachbegriff definieren und von anderen abgrenzen, direkt mit Fachbegriffen beschrieben werden (Inhaltsbeschreibung). Ohne davor eine Situation zu beschrieben. • Ein anderer Fachbegriff wird als gleichbedeutend dargestellt (Ausdrucksersetzung) NICHT: Begriffe Synonym verwenden, ohne zu sagen, dass sie es sind. • Der Fachbegriff und die Einheit werden genannt. NICHT: Zahlenbeispiel: $v = 5\frac{m}{s}$ • NICHT: Formeln nennen. • *Allein durch die Kommentierung muss eindeutig sein, welcher Fachbegriff gemeint ist.*	• Pilot_1_01a 5:20 „Energie ist eine Form von Arbeit“ • Pilot_1_01a 5:58 „Arbeit ... diese Definition ... du setzt eine gewisse Kraft ein, um einen Weg zu absolvieren“ • Pilot_1_06a 4:30 Geschwindigkeit das ist Meter pro Sekunde oder Kilometer pro Stunde • (Siehe: Kommentierung von FB.docx)
Unkommentierte Fachbegriffe	• Alle weiteren Fachbegriffe, die nicht kommentiert werden oder Teil der Kommentierung eines anderen Fachbegriffs sind.	
	Proband → Medium → Verbal → Kontext *Kodieren bei: Vortrag, Fragen und Nonverbal*	
Adressaten in Situation einbeziehen	• Der Adressat wird in die Situation einbezogen. Es wird explizit gesagt, dass der Adressat die Tätigkeit, die im Szenario oder in einem Beispiel beschrieben wird, vollführt. • Du tust..., Wir tun..., Dir/Uns passiert...	• Du fährst in der Achterbahn... • Wenn du eine Kraft ausübst... • Wir sprengen den Asteroiden... (Pilot_2_02a 0:35) • Pilot_1_03a 2:25 • NICHT: „Wenn du etwas beobachtest...“ Pilot_1_03a 1:52

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
(Beispiele)	• *Der Kontext in dem Gesprochen (Vortrag oder Frage) wird, ist nicht das Szenario. Dabei muss ein neuer oder der alte Gegenstand (Motorrad, U-Boot, Asteroid, Achterbahnwagen und Auto) in einer komplett anderen Situation betrachtet werden.*	• *KEIN Beispiel: Fahrrad oder Auto statt Motorrad. Meteorit statt Asteroid.* • *KEIN Beispiel: Die Situation wird nur leicht abgewandelt. z. B. Die Achterbahnschienen hören auf und der Waagen fällt weiter.* • ACHTUNG bei Gegenständen: Wenn explizit der Gegenstand oder „das hier " gesagt wird, ist es ein Beispiel.
alltagsnahe Beispiele	Der Proband zieht Beobachtungen oder körperliche Erfahrungen aus dem Alltag heran, um etwas zu erklären. • Es ist prinzipiell möglich die beschriebene Situation im Alltag (oder im Fernsehen) beobachtet oder selbst erlebt zu haben. • Man kann es während der Erklärung beobachten oder spüren (z. B. in einem Freihandexperiment). • Allgemeinwissen (Dinosaurier, Tsunami auf den Philippinen) • KEIN Beispiel für Erdanziehungskraft: Auf der Erde sein. Sich auf der Erde bewegen.	• „Im Wasser fühlst du dich leichter" Pilot_1_05a 1:15, 1:40 • Wasserflasche mit Wasser gefüllt ist leichter, als eine mit Luft gefüllte Pilot_1_05a 5:00 • „Und zwar sieht man das auch beim Wasserstrahl. Wenn man so einen Schlauch hält und den geradeaus spritzt, dann hat das ungefähr auch so eine Form." Prä_1_08

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
abstrakte Beispiele	Erfahrungen oder Beobachtungen können höchstens aus fachlicher Sicht vorgenommen werden. Im Alltag können sie nicht vorgekommen sein, weil... 1: ... die Beobachtung mit dem bloßen Auge nicht möglich ist (z. B. weil die Bewegung zu schnell ist). 2: ... die Erfahrung ist auf der Erde nicht möglich (z. B. Erfahrungen im Weltall oder auf dem Mond) 3: ... eine Situation oder eine Gegenstand abstrakt beschrieben werden und daher keinen Alltagskontext haben. 4: ... die Situation absurd ist. Sie wird nur für die Erklärung konstruiert. 5: ... es in der Situation nicht um Beobachtungen oder Erfahrungen geht.	1: Beim Fallen wird der Stift immer schneller. Flugbahn des Stiftes ist eine Parabel. 2: „Stell dir vor, du müsstest ein 60 km großes Auto zur Seite schieben“ Pilot_2_02a 4:05 3: Etwas an einer Schnur im Kreis schleudern (0:31) und die schiefe Ebene. Pilot_1_01b 2:58 4: Kurz bevor man gegen eine Wand fährt, Kugeln aus dem Auto werfen. Pilot_1_06a 3:46; Mit einer Bowlingkugel auf dem Dach einen Unfall haben. Pilot_3_03b 5: Arbeit, wenn eine Kiste gezogen wird (6:20) und eine Kiste die Treppe hoch tragen wird. Pilot_1_01a 7:00

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
unangemessene Beispiele	1: Eine Situation wird falsch beschrieben. 2: Die Situation passt nicht zum physikalischen Inhalt.	1: Pilot_1_06a 3:22 Aus der Perspektive, aus der das Beispiel beschreiben wird, sieht die Flugbahn der Kugel anders aus. 2: Pilot_1_01a 1:50 Tisch umschubsen , um daran den Impuls zu erklären 2: Prä_2_19 P: „Sagen wir mal meine Hand. Die läuft immer weiter, immer weiter (P. bewegt die Hand entlang einer Geraden) und damit ich diese meine Hand eine Kurve macht, brauch ich eine Kraft, die in die Mitte zieht. Das heißt, in diesem Fall spann ich jetzt meinen Arm und ziehe halt (P. bewegt seine Hand im Halbkreis).“
ohne Kontext	• Der Proband erklärt unabhängig von einem Kontext (weder das Szenario noch ein Beispiel). Der Proband spricht davon, dass alles bzw. irgendetwas sich so verhält. Sagt Material, Objekt, Gegenstände... • Der Proband trägt Definitionen vor.	• „Es gibt in der Mechanik sogenannte Actio- und Reactiogesetze. Das heißt, wenn du auf etwas eine Kraft ausübst, wirkt es auch eine Kraft auf dich aus.“ Pilot_1_1a 1:20 • „Arbeit ist letztendlich die zurückgelegte Kraft über einen Weg“ Pilot_1_1a 1:20 5:40
verknüpfen von Beispielen	• Der Proband stellt eine Verknüpfung zwischen zwei Beispielen her. Verknüpfung geschieht indem er für das Beispiel spezifische Begriffe in dem anderen Beispiel verwendet oder explizit sagt, was gleich ist. • Es können auch alltagsnahe Beispiele mit unangemessenen verknüpft werden oder kontextlose Erklärungen mit einem Beispiel.	• Hier ist das genauso

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
verknüpfen von Kontexten	• Der Proband stellt explizit eine Verknüpfung zwischen zwei Kontexten her. (Szenario - Beispiel; Szenario - ohne Kontext, Beispiel - Beispiel etc.) • Verknüpfung geschieht durch gleiche Wortwahl oder indem explizit gesagt wird, was gleich ist. • Eine Abbildung auf der ein Kontext abgebildet ist, wird genutzt, um einen anderen zu erklären. • Die Verknüpfung muss sinnvoll sein. D.h. bei den Kontexten ist der Aspekt wirklich analog. (Gegenbeispiel: 1_3_14 7:00 Apfel am Baum und Reibung in der Kurve)	• „Der Asteroid bekommt einen Stoß.“ (Wortwahl aus dem Beispiel mit Billardkugeln. In dem Fall muss auch die Kategorie „alltagsnahes Beispiel“ markiert werden.) • „Hier ist das genauso.“ aber nur mit einer expliziten Erklärung, was genauso ist. Das genauso darf nicht einfach so auf die gesamte Situation bezogen sein.
	Proband → Mathematisierungen *Mathematische Sprache - Verwendung einer besonderen Form von Fachsprache: Mathematische Begriffe, Abkürzungen, Formelzeichen und Zahlen*	
Zahlenbeispiel	• Der Proband ersetzt eine physikalische Größe durch eine Zahl oder nennt (bzw. notiert) sie als Beispiel für eine Größe. • Eine Zahl kommt in einer Formel vor: Zahlenbeispiel (NICHT mehr unter Formel kodieren). • NICHT: Anzahlen, Stückzahlen, z. B. „in zwei Teile“ oder „die Hälfte„ oder ”doppelt‘“ • NICHT Besondere Eigenschaften: „μ liegt zwischen 0 und 1“	• Nach einer bestimmten Zeit, sagen wir mal zwei Sekunden. • Hier wäre sie [die Strecke] gerade v_0 mal eine Sekunde. • Pilot_2_ Ein 60 km großes Auto • Die Erdbeschleunigung g ist $9,81\frac{m}{s^2}$ • Pilot_3_02b 5:32 „Wenn ich jetzt einen Kilometer pro Stunde fahre, dann fahre ich in einer Stunde einen Kilometer“

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Formel (mit Rechen-zeichen)	• Der Proband nennt (oder notiert) eine Formel in der ein Rechenzeichen explizit vorkommt. (plus, minus, mal, geteilt, pro, durch, Produkt, addieren...) • NICHT: zusammengesetzte Einheit z. B. m/s	• Die Formel für die Strecke ist $s = v \cdot t$. • „Das Produkt aus v und t...“ oder nur „Das Produkt“, „Die Summe“ • „... Dichte, also Masse pro Volumen“ Pilot_1_08a 4:41 • Formel aufschreiben, ohne sie zu nennen. Pilot_2_3b 1:22
Je-Desto-Zusam-menhänge	• Der Proband beschreibt Proportionalitäten, indem er ”je-desto-Aussagen “ macht.	• Je schneller er über den Hügel fährt, desto breiter muss der Hügel sein. • Wir müssen bei konstanter Geschwindigkeit die Zeit erhöhen, damit die Strecke größer wird. • NICHT: Keine mathematische Funktion oder Proportionalität. z. B. Je größer die Knautschzone, desto besser für den Insassen. Pilot_2_03a

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Mathematische Begriffe/ Idealisierungen	• Der Proband nennt einen mathematischen Begriff. • Der Proband macht eine Idealisierung bei der ein mathematischer Begriff verwendet wird. (Parabel, Kreis, konstant)	• Volumen, Höhe/Teiefe/Breite, Kugel (wenn es dabei um die Form geht) • Vektor, Satz des Pythagoras, Steigung • Kreis, Kreisradius, Mittelpunkt, rechter Winkel • Parabel, Funktion, Tangente, Steigung, Hochpunkt, Gerade, Steigungsdreieck • Koordinatensystem, Achsenkreuz, x- bzw. y-Richtung, nach x bzw. y, Minusbereich und Plusbereich, • Vektor, vektorielle Größe • konstant, Konstante (Prä_1_08 9:50 „v_x ist konstant.“) • NICHT Prä_1_08 4:44 „Der hätte hier in die Richtung immer die gleiche Geschwindigkeit“ - weil nur eine Idealisierung gemacht wird, ohne mathematische Begriffe zu verwenden

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
	Proband → Medium → Verbal → Fragen stellen *Explizit formulierte Fragen, die der Proband an den Adressaten stellt. D. h. ein Fragewort ist enthalten, der Satzbau entspricht einer Frage oder eine Aussage mit einer Annahme über den Adressaten wird gemacht. z. B. Das hast du bestimmt in der Schule gemacht. (Diese muss vom Adressaten beantwortet werden) NICHT: Eine Erklärung wird geben und am Ende des Satzes wird die Stimme gehoben oder wird ein „ne?, oder?, stimmts?, richtig?...“ ran gesetzt. Das ist nur Grounding. NICHT: rhetorische Fragen, ohne eine Antwort abzuwarten*	
Vorhersage	• Der Proband stellt mindesten eine geeignete Frage, damit der Adressat den betreffenden Sachverhalt selber herausfindet und beachtet dabei die beiden folgenden Schritte: 1) Die Vorstellungskraft des Adressaten wird angeregt, indem eine Situation beschrieben wird. 2) Der Proband fragt nach einer Vorhersage für diese Situation. • Fragen im Sinne von „Was passiert, wenn...?“ Dabei geht es nur um die Situation und nicht um physikalische Größen. • NICHT: P. fragt direkt nach der Erklärung des Szenarios. „Ich will dir erklären,...warum nach Regenfällen so häufig Unfälle passieren. So, was meinst du denn?“ (Prä_1_Katja, Igor oder Lisa	• Prä_1_08 1:14 „Stell dir vor, jetzt würden hier die Schienen aufhören. Was glaubst du, was würde passieren?“ und dann bei 1:30 „Hast du auch eine Idee, wie der runter fallen würde?“

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Verständnisversicherung (direkt)	• Proband erfragt Erfolg seiner Erklärung, also ob der Proband ihn verstanden hat oder ihm zustimmt. Wird eingestuft, wenn deutliche Frage dementsprechend erfolgt. • Die Frage hat keinen Aufforderungscharakter (und bei der Antwort gibt es kein richtig oder falsch, weil nach der Sicht des A. gefragt wird). • Explizite Frage oder Äußerung die darauf abzielt, ob der A. es verstanden hat, sich das vorstellen kann, etwas genauso sieht oder was noch erklärt werden soll.	• „Das kannst du dir vorstellen?“ Pilot_1_01a 2:19 • „Kannst du dir das vorstellen?“ Pilot_1_01a 3:00 • Pilot_1_05a 8:00 • „Kennst du...„ oder „Weist du...“ • „Stell noch eine Frage“ Pilot_3_02b 4:17 • „Hast du noch irgendwelche Fragen?“ (Verständnis bezogen auf die das ganze Szenario, statt auf die Erklärung genau davor.) • „wie möchtest du es erklärt haben? Was genau verstehst du nicht?“ Pilot_1_03a 8:11 • „Jetzt habe ich dich verwirrt“ 1_2_06 6:05
Verständnisversicherung (diagnostizierend)	• Proband erfragt Erfolg seiner Erklärung, indem er ihm/ihr eine Wissensfrage oder Aufgabe stellt. Diese Aufgabe bezieht sich auf etwas, was schon erklärt wurde. (z.B. soll der Sachverhalt auf einen neuen oder leicht abgeänderten Kontext bezogen werden) • Je nach dem, ob die Frage richtig oder falsch beantwortet wurde, kann der Proband auf das Verständnis schließen.	• „Wenn Luft hinzukommt und Wasser raus, was wird dann passieren?“ Pilot_1_08a 7:35 (vorher erklärt er, dass wenn Wasser in das U-Boot reinkommt, der Auftrieb kleiner wird und das U-Boot sinkt) • „Jetzt teilt der Asteroid sich in vier Teile. Nehmen wir mal dieses Teilchen. Wie würde es sich weiterbewegen?“ Prä_1_02a 9:40 • „Könntest du das so einigermaßen wiedergeben, wie du das verstanden hast„ Pilot_1_09a 7:53

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Frage nach Vorerfahrungen/ -wissen	• • Proband erfragt Alltagserfahrungen oder Alltagswissen des Adressaten. • Adressat erfragt Wissen über die Erdklärsituation (z. B. ob die Aufgabenstellung bekannt ist.) • KEIN physikalisches Wissen. Es geht nur um Situationen und Beschreibungen derer.	• Warst du schon mal tauchen? Bist du schon mal Achterbahn gefahren? • Kennst du Crash Tests? • Was weißt du über Asteroiden, bzw. Knautschzonen, bzw. Unfälle in Kurven? • Kennst du die Aufgabe? • „Was kann man (noch) machen, um zu verhindern, dass der Asteroid auf die Erde trifft?“ Prä_1_2a 0:12 und Prä_2_05b 1:08
Vorwissensabfrage (direkt)	• Proband erfragt schulisches (mathematisches und naturwissenschaftliches) Vorwissen des Adressaten mit einer expliziten Frage. • Die Frage hat keinen Aufforderungscharakter und es gibt keine richtige oder falsche Antwort auf die Frage. • Eine Annahme über das Wissen oder den Unterricht des Adressaten wird explizit genannt. • NICHT: Eine Aussage zu einem physikalischen Sachverhalt wird gemacht und am Ende wird die Stimme gehoben, z. B. Das ist wegen der newtonschen Axiome?	• „Newtonsche Axiome?“ Pilot_2_03b 8:46 (- weil das nicht mit einer Aussage verbunden ist.) • Annahmen über das Wissen des Adressaten: „Das kennst du bestimmt aus deinem Physikunterricht?“ Pilot _1_05a 0:48 oder „Das habt ihr in der Schule eventuell schon gemacht... “ Prä_1_09 10:03 • „Kennst du...?“ oder „Weist du...?“ oder „Hattet ihr ... schon in der Schule?“

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Vorwissensabfrage (diagnostizierend)	• Proband erfragt Vorwissen des Adressaten indirekt, indem er ihm eine Aufgabe stellt. • Die Frage bezieht sich auf einen physikalische Sachverhalt oder es soll physikalisch erläutert werden, WARUM eine Situation so ist. • Stellt eine Wissensfrage zu einem physikalische Thema. • Die Frage hat einen Aufforderungscharakter, also wird vom Adressaten mehr erwartet, als nur mit Ja und Nein zu antworten. • Die Antwort auf die Frage kann richtig oder falsch sein und daraus kann der Proband darauf schließen, ob das Vorwissen beim Adressaten vorhanden ist oder nicht.	• „Wenn du dahin fahren willst, wohin muss dann die Kraft einmal wirken?“ • „Hast du ein Vorstellung/Idee, WARUM das Motorrad aus der Kurve fliegt, bzw. WARUM die Asteroiden so weiterfliegen, bzw. Autos mit Knautschzonen sicherer sind?“
weitere Fragen	• Es werden Fragen gestellt, die nicht in die obigen Kategorien eingeordnet werden können. • Z. B. Rückfragen zu Adressatenäußerungen oder was auf dem Bild zu sehen ist. • Fragt, was auf einer Abbildung zu sehen ist. Lässt eine Abbildung beschreiben. • Fragen, die sich darauf beziehen die Situation zu beschreiben (also nicht mit physikalischen Begriffen zu erklären).	• Prä_1_08 4:40 „Meinst du so was hier?“ • Proband legt ein Bild hin: „Kannst du mal beschreiben, was das Auto macht?“ Pilot_1_04a 0:24 • Schreibt ihr eine Klausur darüber? • Bekommst du das noch von anderen erklärt?
	Proband → Medium → Verbal → Handlungsanweisungen geben *Unabhängig von Fragen kodieren, d. h. eine Äußerung kann Frage und Handlungsanweisung zugleich sein.*	

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Handlungs-aufforderung	• Der Proband fordert den Adressaten auf etwas zu tun. • z.B. etwas aufzuzeichnen, etwas in eine Skizze zu zeichnen, etwas mit Gegenständen durchzuführen. • NICHT: Guck dir das mal an. Lies dir das mal durch.	• Prä_1_08 „Zeichne mal ein, wie du das meinst“ • Pilot_1_Igor
	Proband→ Medium → Nonverbal	
Abbildungen nutzen - allgemein	• *Kodieren, wenn Vortrag oder Fragen kodiert wird!* • *Bei Vortrag kodieren, wenn die Abbildung beschrieben wird oder etwas anhand der Abbildung erklärt wird.* • *Bei Fragen kodieren, wenn gefragt wird, was auf dem Bild zu sehen ist oder wenn das Bild genutzt wird um die Frage zu formulieren (also auch kurze Verweise).* • *Es können auch mehrere Abbildungen gleichzeitig zu einem Erklärungssegment gehören, wenn über beide gleichzeitig oder im Wechsel gesprochen wird. Z. B., wenn zwei Abbildungen miteinander verglichen werden.* • *NICHT: Verweis auf eine Abbildung während des Vortragens (Kurzes zeigen auf die gesamte Abbildung oder ein einzelnes Element in einer Abbildung)* • *: NICHT: Abbildung ins Blickfeld legen und nichts dazu sagen.*	*Beispiele für Verweise auf andere Abbildungen:* • *Prä_08 2:17-2:30 erklärt, was auf der Abbildung zu sehen ist mit einem Verweis auf eine andere Abbildung bei 2:18 („das auch“)* • *Prä_08 2:55-3:15 erklärt den Verlauf der Achterbahnschienen mit einem Verweis bei 3:08* • *Prä_08 6:25-6:45 erklärt an einer Abbildung mit einem Verweis auf eine andere Abbildung bei 6:29*

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
reales Bild (Ebene 1)	• Der Proband erklärt anhand eines realen Bildes, also einer Abbildung die nur reale Gegenstände zeigt ohne physikalische Elemente. • Zeichnungen und Erklärungshilfekarten mit den Nummern 1a, 1b oder 1c. • Zeichnungen die erstellt werden: Endzustand der Zeichnung in die Erklärungsebenen einordnen. Also das Aussehen der Zeichnung, wenn ein Segmentwechsel weg von der Zeichnung kommt.	
Bildanalogie (Ebene 2)	• Der Proband erklärt anhand einer Bildanalogie, also eines Bildes das Elemente eines realen Bildes und eines logischen Bildes enthält. • Zeichnung und Erklärungshilfekarten mit den Nummern 2a, 2b oder 2c. • Zeichnungen die erstellt werden: Endzustand der Zeichnung in die Erklärungsebenen einordnen. Also das Aussehen der Zeichnung, wenn ein Segmentwechsel weg von der Zeichnung kommt.	
logisches Bild (Ebene 3)	• Der Proband erklärt anhand eines logischen Bildes, also eines Bildes das keine realen Elemente enthält. • Zeichnung und Erklärungshilfekarten mit den Nummern 3a, 3b oder 3c. • Zeichnungen die erstellt werden: Endzustand der Zeichnung in die Erklärungsebenen einordnen. Also das Aussehen der Zeichnung, wenn ein Segmentwechsel weg von der Zeichnung kommt.	

26

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Gegenständliche Hilfsmittel (zeigen)	Gegenstände (Stift, Glas, Tisch, Papier, Hand etc.) werden zum erklären genutzt. Mit den Gegenständen wird etwas gezeigt, vergleichbar mit einem Bild oder ein kleines Experiment durchgeführt. Der Inhalt wird durch die Handlung transportiert und mit der verbalen Sprache wird diese nur kommentiert. Die eigentliche Erklärung liegt im gezeigten, also das Gesagte unterstützt das Gezeigte und nicht umgekehrt. Dies ist erfüllt, wenn eines der folgenden Dinge zutrifft: 1: Der Gegenstand verlässt die Hand 2: Mit den Gegenständen oder Bewegungen wird eine Situation nachgestellt. 3: Ein Bild (ähnlich wie eine Abbildung) wird mit Hilfe der Gegenstände gezeigt. 4: Eine Bewegung wird in mehrere einzelne Schritte unterteilt. 5: Zwei Formen oder Richtungen werden mit zwei Bewegungen oder zwei Gegenständen dargestellt und miteinander verknüpft. • Nur bei Beispiele kodieren, wenn es verbal ein Beispiel ist.	1: Glas wird über den Tisch geschubst Prä_1_10, 06:01-06:26, Stift wird geworfen Pilot_1_08b 1:18-1:38 2: Pilot_1_08a 5:33-7:00 Ein zugehaltenes Glas ist das U-Boot 3: Bild mit Gegenständen zeigen Pilot_1_08b 3:11 4: Einzelne Bewegungsschritte Pilot_1_08b 7:54 5: Verknüpfung von zwei Formen oder Richtungen Pilot_1_08b 4:50
experimentieren	• Der Punkt 1 trifft zu: „Der Gegenstand verlässt die Hand“ oder der Adressat nimmt etwas in die Hand bzw. tut etwas und kann dabei etwas spüren.	• Glas wird über den Tisch geschubst Prä_1_10, 06:01-06:26, Stift wird geworfen Pilot_1_08b 1:18-1:38 • A. nimmt Buch in die Hand Prä_09. oder A. streicht mit der Hand über den Tisch.

Kapitel 1 Kodiermanual

Kategorie	Beschreibung / Indikator	Ankerbeispiel
Handlung/ Gegenstände mit Szenario verknüpfen	• Benennt den Gegenstand mit einem Objekt aus dem Szenario.	• Proband nimmt ein Buch und ein Saftpäckchen und sagt„Das bist du... Das ist der Sitz...“ Prä_1_09a 2:40 Im weiteren Szenario: Zeigt extra auf das Buch und sagt „der Sitz“
Schrift (ohne Beschriftungen)	• Der Proband schreibt Stichpunkte, Sätze, Formeln oder Formelzeichen auf oder zeigt etwas an der Schrift. • NICHT: Beschriftung einer Zeichnung. Das wird bei der Zeichnung kodiert.	•
verknüpfen von nonverbalen Elementen	• Proband sagt explizit (also auf verbaler Ebene), was an zwei nonverbalen Elementen gleich ist (z. B. Wo ein Aspekt in dem anderen Bild wiederzufinden ist, oder welche generellen Eigenschaften die Bilder verbindet). • Proband wechselt andauernd zwischen zwei nonverbalen Elementen, während er etwas erklärt. (Wichtig: Gleichzeitiges oder abwechselndes betrachten der Abbildungen und nicht nacheinander.) • Alle nonverbalen Elemente können untereinander verknüpft werden, also z. B. zwei Abbildungen gleicher oder unterschiedlicher Ebenen oder Abbildung mit Gegenstand oder mit Schrift. • NICHT: Zwei Abbildungen hinlegen und sagen „Das ist das gleiche“ und sonst nichts mehr.	• Zwei Abbildungen: Prä_1_08 6:15- 6:25 Verknüpfung über die Wortwahl „...genauso wie...“ (Auch bei 6:48-7:09 und 7:19 - 7:28, 7:12) • Zwei Abbildungen: Pilot_1_9a 3:37 Erklärt, wo der Achterbahnwagen und die Schienen aus dem einem Bild im anderen wären. • Zwei Abbildungen: Pilot_1_10a 0:41- 0:52 Sagt, dass die Kurve in dem einen Bild als der Kreis in dem anderen Bild betrachtet werden kann. • Gegenstand und Abbildung: Prä_1_10 „Das ist genauso hier...“ Das Glas soll so an geschubst werden, dass es die gleiche Kurvenbahn fährt, wie auf der Erklärungshilfekarte zu sehen ist. • Schrift und Abbildung: Prä_1_08 10:31 Zeigt abwechselnd auf die Formel und die Abbildung.

1.2 Kategorien Erklärer

Kategorie	Beschreibung / Indikator	Ankerbeispiel
erstellen/ verändern	• der Proband erstellt oder verändert eine Abbildung oder etwas Geschriebenes. Der Proband schreibt, zeichnet oder verändert die Abbildung durch Abdecken von Teilen.	• Ein Teil der Abbildung wird zugedeckt. Pilot_1_01b 5:19
	Proband → Medium → Verbal → Strukturierung der Erklärung Teil 2	
Unterbrechung	• Proband unterbricht Adressaten, wenn der etwas darlegt. Wird eingestuft, wenn die Unterbrechung einen Gedankengang des Adressaten abbricht.	-
Antwort unangemessen/ungenügend	• Proband kann oder will eine direkte Frage nicht oder nicht angemessen beantworten. • Proband kann oder will auf einen Impuls des Adressaten nicht eingehen. • „Überhört" die Frage oder den Impuls und reagiert überhaupt nicht darauf. • Die vermeintliche Antwort bezieht sich nicht auf die Frage. Z.B. wird gefragt, was kinetische Energie ist, aber in der Antwort wird auf die Energieerhaltung eingegangen. • NICHT: fachlich falsche Antworten	• Kurvenfahrt: A: „Ich muss so was bestimmt berechnen in der Klausur" P: „Nein, ich glaube eher, dass der freie Fall vorkommt" • A: „Schwerelosigkeitsgefühl, ist das das gleiche wie im Weltall?" P: „Das kann ich dir nicht sagen, weil ich noch nie im Weltall war." Pilot_ 1_03a 0:30 • A: „Gibt es die Abbildung noch einfacher?" P: „Nein, die gibt es nicht viel einfacher." (ohne zu versuchen der Adressatin das Thema oder die Abbildung nochmal zu erklären) Pilot_1_02a 3:58 • Adressatin fragt nach einem Fachbegriff und der Proband verwendet einen anderen in der Erklärung. Pilot_1_04b 4:41

	Oberkategorie	Kategorie	Erläuterung
Variablen einer Erklärung	Sprache	1. Umschreibung von Fachbegriffen	1. Meint die direkte Erläuterung eines Fachbegriffs in Alltagssprache.
	Beispiele/ Analogien	2. Unangemessenes Beispiel	2. Mein Beispiele, die ein falsches Bild der Realität vermitteln.
	Mathemati-sierung	3. Zahlenbeispiel	3. Meint z.B. physikalische Größen, die durch einen Wert veranschaulicht werden.
	Graphische Darstellungs-form	4. Verknüpfen von nonverbalen Elementen 5. Gegenständliche Hilfsmittel verwenden 6. Gegenstände mit Erklärung verknüpfen 7. Ad-hoc Experimente	4. Meint z.B. das Aufeinanderbeziehen verschiedener Graphiken („Hier sieht man... und da sieht man..."). 5. Mein das Einbeziehen von Gegenständen (z.B. Wasserflasche) in die Erklärung zur Nachstellung von Situationen oder Bewegungen. 6. Meint das Einbeziehen von Gegenständen durch Benennung, was diese repräsentieren („Das bist du..."). 7. Meint das spontane Durchführen von „Experimenten", z.B. Werfen eines Papierknäuels zur Veranschaulichung der Wurfparabel.
Perspektive	Adressaten-gemäßheit diagnostizieren	8. Diagnostizierende Verständnisversicherung 9. Aufforderung des Adressaten zur Handlung	8. Meint das Diagnostizieren des Erklärungserfolgs durch Wissensfragen oder Aufgaben – NICHT: „Hast du das verstanden?". 9. Meint die Aufforderung, eine Handlung wie z.B. zeichnen, zu vollführen.
Verknüpfung von Erklärungen	Strukturelle Elemente	10. Strukturierender Rückblick 11. Zusammenfassung/ Wiederholung 12. Ermunterung und Lob	10. Meint, dass der Erklärer explizit darauf verweist, dass etwas bereits erklärt wurde und jetzt eine Rolle spielt („Du hast ja gerade von der Zentrifugalkraft gesprochen..."). 11. Meinet das Zusammenfassen eines bereits erklärten Sachverhalts. 12. Meint Lob bei erfolgreichem Handeln des Schülers oder Verständnis für etwaige Schwierigkeiten.

Tabelle 1: Liste der Kategorien, die in das Maß eingehen sowie deren Zugehörigkeit zu den sich aus dem Modell ergebenen Oberkategorien

Pretest

Kepler'sche Gesetze

Bitte gib die ersten zwei Buchstaben des Vornamens deiner Mutter ein:		
Bitte gib die ersten zwei Buchstaben deines Vornamens ein:		
Bitte gib die ersten zwei Ziffern deines Geburtstages ein (zweistellig):		

Pro Frage ist immer nur **eine** Antwort korrekt!

1) Worauf beruhen die Kepler'schen Gesetze?

a) Auf Newtons Gravitationsgesetzen

b) Auf reinen Gedankenexperimenten

c) Auf Beobachtungen von Planetenbahnen und daraus resultierende Messungen

d) Auf Übereinstimmungen von Wissenschaftlern, die auf Kongressen erarbeitet wurden.

a) ☐ b) ☐ c) ☐ d) ☐

2) Was lässt sich mit Hilfe der Kepler'schen Gesetzen beschreiben?

a) Sie beschreiben die zeitlichen und räumlichen Bewegungsänderungen der Sonne von der Erde aus betrachtet.

b) Sie beschreiben sowohl die zeitlichen als auch die räumlichen Abläufe der Planetenbewegungen um die Sonne.

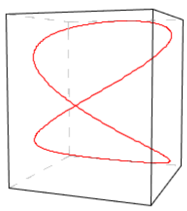
c) Sie beschreiben sowohl die zeitlichen als auch die räumlichen Abläufe von Gegenständen auf Kreisbahnen.

d) Sie beschreiben nur die Bewegungsbahnen von Asteroiden.

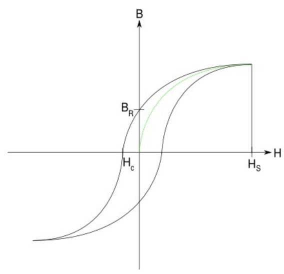
a) ☐ b) ☐ c) ☐ d) ☐

3) Welche Abbildung zeigt eine Ellipse?

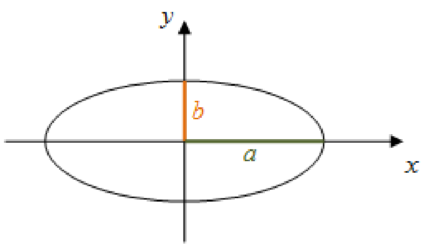
a)



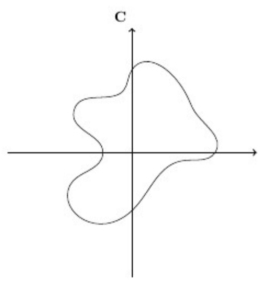
b)



c)



d)



a) ☐ b) ☐ c) ☐ d) ☐

4) Wie viele Brennpunkte hat eine Ellipse <u>höchstens</u>?

a) 0 (null) — a) □

b) 1 (einen) — b) □

c) 2 (zwei) — c) □

d) 3 (drei) — d) □

5) Wie lautet das <u>zweite</u> Kepler'sche Gesetz?

a) Jeder Körper verharrt im Zustand seiner der Ruhe oder gleichförmigen Bewegung, sofern er nicht durch eine Kraft gezwungen wird, seinen Bewegungszustand zu ändern. — a) □

b) Die von der Sonne zu einem Planeten gezogene Verbindungsgerade überstreicht in gleichen Zeiten gleiche Flächen. — b) □

c) Alle Planeten bewegen sich auf Ellipsenbahnen, in deren einem Brennpunkt die Sonne steht. — c) □

d) $C = 4\pi (\gamma M_{\Theta})^{-1}$, mit γ = Gravitationskonstante und M_{Θ} = Masse der Sonne — d) □

6) Wie ist ein Fahrstrahl definiert?

a) Ein Fahrstrahl beschreibt eine Verbindungslinie zwischen dem Mittelpunkt (genauer Schwerpunkt) eines Planeten und dem Gravitationszentrum um das sich der Planet bewegt (oft zur Vereinfachung Sonnenmittelpunkt). — a) □

b) Ein Fahrstrahl ist die Strecke, die ein Planet auf seiner Umlaufbahn im Laufe eines Erdenjahres zurückgelegt hat. — b) □

c) Ein Fahrstrahl beschreibt die exakte Umlaufbahn von Planeten, Satelliten, Raketen und anderen Himmelskörpern. — c) □

d) Ein Fahrstrahl beschreibt die gedachte direkte Linie zwischen zwei Umlaufbahnen zweier Planeten. — d) □

7) Aus dem <u>zweiten</u> Kepler'schen Gesetz lässt sich eine wichtige Folgerung ableiten.

Vervollständige dazu folgende Aussage:

Je näher sich ein Planet auf seiner Orbitalbahn der Sonne nährt, desto...

a) ... langsamer wird er. — a) □

b) … weiter entfernt sind die anderen Planeten. — b) □

c) … schneller wird er. — c) □

d) … schneller rotiert er um seine eigene Achse. — d) □

Posttest

Kepler'sche Gesetze

Bitte gib die ersten zwei Buchstaben des Vornamens deiner Mutter ein:		
Bitte gib die ersten zwei Buchstaben deines Vornamens ein:		
Bitte gib die ersten zwei Ziffern deines Geburtstages ein (zweistellig):		

Pro Frage ist immer nur **eine** Antwort korrekt!

1) Worauf beruhen die Kepler'schen Gesetze?

a) Auf Newtons Gravitationsgesetzen

b) Auf reinen Gedankenexperimenten

c) Auf Beobachtungen von Planetenbahnen und daraus resultierende Messungen

d) Auf Übereinstimmungen von Wissenschaftlern, die auf Kongressen erarbeitet wurden.

a) □ b) □ c) □ d) □

2) Was lässt sich mit Hilfe der Kepler'schen Gesetzen beschreiben?

a) Sie beschreiben die zeitlichen und räumlichen Bewegungsänderungen der Sonne von der Erde aus betrachtet.

b) Sie beschreiben sowohl die zeitlichen als auch die räumlichen Abläufe der Planetenbewegungen um die Sonne.

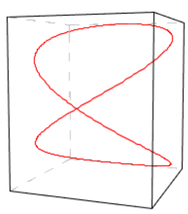
c) Sie beschreiben sowohl die zeitlichen als auch die räumlichen Abläufe von Gegenständen auf Kreisbahnen.

d) Sie beschreiben nur die Bewegungsbahnen von Asteroiden.

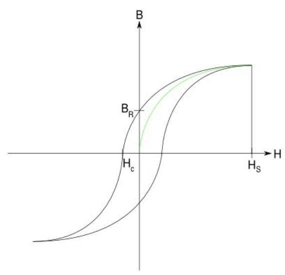
a) □ b) □ c) □ d) □

3) Welche Abbildung zeigt eine Ellipse?

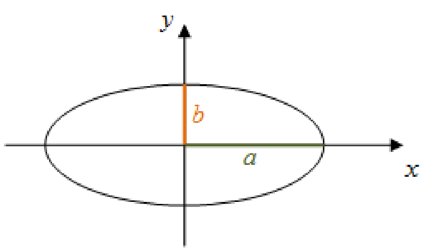
a)



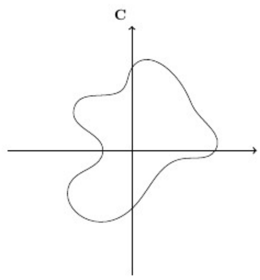
b)



c)



d)



a) □ b) □ c) □ d) □

4) Wie viele Brennpunkte hat eine Ellipse <u>höchstens</u>?

a) 0 (null) a) □

b) 1 (einen) b) □

c) 2 (zwei) c) □

d) 3 (drei) d) □

5) Wie lautet das <u>zweite</u> Kepler'sche Gesetz?

a) Jeder Körper verharrt im Zustand seiner der Ruhe oder gleichförmigen Bewegung, sofern er nicht durch eine Kraft gezwungen wird, seinen Bewegungszustand zu ändern. a) □

b) Die von der Sonne zu einem Planeten gezogene Verbindungsgerade überstreicht in gleichen Zeiten gleiche Flächen. b) □

c) Alle Planeten bewegen sich auf Ellipsenbahnen, in deren einem Brennpunkt die Sonne steht. c) □

d) $C = 4\pi (\gamma M_{\Theta})^{-1}$, mit γ = Gravitationskonstante und M_{Θ} = Masse der Sonne d) □

6) Wie ist ein Fahrstrahl definiert?

a) Ein Fahrstrahl beschreibt eine Verbindungslinie zwischen dem Mittelpunkt (genauer Schwerpunkt) eines Planeten und dem Gravitationszentrum um das sich der Planet bewegt (oft zur Vereinfachung Sonnenmittelpunkt). a) □

b) Ein Fahrstrahl ist die Strecke, die ein Planet auf seiner Umlaufbahn im Laufe eines Erdenjahres zurückgelegt hat. b) □

c) Ein Fahrstrahl beschreibt die exakte Umlaufbahn von Planeten, Satelliten, Raketen und anderen Himmelskörpern. c) □

d) Ein Fahrstrahl beschreibt die gedachte direkte Linie zwischen zwei Umlaufbahnen zweier Planeten. d) □

7) Aus dem <u>zweiten</u> Kepler'schen Gesetz lässt sich eine wichtige Folgerung ableiten.

Vervollständige dazu folgende Aussage:

Je näher sich ein Planet auf seiner Orbitalbahn der Sonne nährt, desto...

a) ... langsamer wird er. a) □

b) … weiter entfernt sind die anderen Planeten. b) □

c) … schneller wird er. c) □

d) … schneller rotiert er um seine eigene Achse. d) □

11) Wie sehr hat dir das Video allgemein gefallen?

☐	☐	☐	☐	☐
Sehr gut	gut	mittel	schlecht	sehr schlecht

12) Hast du davor schon Erklärvideos (z.B. von YouTube, Sofatutor, Draufhaber, etc.) genutzt?

☐	☐	☐	☐	☐
sehr häufig	oft	manchmal	selten	nie

13) Kannst du dir vorstellen, das Video als Vorbereitung für Klausuren, Teste oder Referate zum Thema „Kepler'sche Gesetze“ zu nutzen?

☐	☐	☐	☐	☐
auf jeden Fall	eher schon	vielleicht	eher nicht	gar nicht

14) Hast du durch das Video Lust bekommen, dich weiter mit Kepler und den nach ihm benannten Gesetzen zu beschäftigen?

☐	☐	☐	☐	☐
Sehr viel Lust	ein bisschen Lust	weiß nicht	wenig Lust	gar keine Lust

15) Wie würdest du die Erklärqualität dieses Videos einschätzen?

☐	☐	☐	☐	☐
Sehr gut	gut	mittel	schlecht	sehr schlecht

Upper Critical Values of Spearman's Rank Correlation Coefficient R_s

Note: In the table below, the critical values give significance levels as close as possible to but not exceeding the nominal α.

	Nominal α					
n	0.10	0.05	0.025	0.01	0.005	0.001
4	1.000	1.000	-	-	-	-
5	0.800	0.900	1.000	1.000	-	-
6	0.657	0.829	0.886	0.943	1.000	-
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.678	0.727	0.818
13	0.385	0.484	0.560	0.648	0.703	0.791
14	0.367	0.464	0.538	0.626	0.679	0.771
15	0.354	0.446	0.521	0.604	0.654	0.750
16	0.341	0.429	0.503	0.582	0.635	0.729
17	0.328	0.414	0.488	0.566	0.618	0.711
18	0.317	0.401	0.472	0.550	0.600	0.692
19	0.309	0.391	0.460	0.535	0.584	0.675
20	0.299	0.380	0.447	0.522	0.570	0.662
21	0.292	0.370	0.436	0.509	0.556	0.647
22	0.284	0.361	0.425	0.497	0.544	0.633
23	0.278	0.353	0.416	0.486	0.532	0.621
24	0.271	0.344	0.407	0.476	0.521	0.609
25	0.265	0.337	0.398	0.466	0.511	0.597
26	0.259	0.331	0.390	0.457	0.501	0.586
27	0.255	0.324	0.383	0.449	0.492	0.576
28	0.250	0.318	0.375	0.441	0.483	0.567
29	0.245	0.312	0.368	0.433	0.475	0.558

(Continued)

	Nominal α					
30	0.240	0.306	0.362	0.425	0.467	0.549
31	0.236	0.301	0.356	0.419	0.459	0.540
32	0.232	0.296	0.350	0.412	0.452	0.532
33	0.229	0.291	0.345	0.405	0.446	0.525
34	0.225	0.287	0.340	0.400	0.439	0.517
35	0.222	0.283	0.335	0.394	0.433	0.510
36	0.219	0.279	0.330	0.388	0.427	0.503
37	0.215	0.275	0.325	0.383	0.421	0.497
38	0.212	0.271	0.321	0.378	0.415	0.491
39	0.210	0.267	0.317	0.373	0.410	0.485
40	0.207	0.264	0.313	0.368	0.405	0.479
41	0.204	0.261	0.309	0.364	0.400	0.473
42	0.202	0.257	0.305	0.359	0.396	0.468
43	0.199	0.254	0.301	0.355	0.391	0.462
44	0.197	0.251	0.298	0.351	0.386	0.457
45	0.194	0.248	0.294	0.347	0.382	0.452
46	0.192	0.246	0.291	0.343	0.378	0.448
47	0.190	0.243	0.288	0.340	0.374	0.443
48	0.188	0.240	0.285	0.336	0.370	0.439
49	0.186	0.238	0.282	0.333	0.366	0.434
50	0.184	0.235	0.279	0.329	0.363	0.430
51	0.182	0.233	0.276	0.326	0.359	0.426
52	0.180	0.231	0.274	0.323	0.356	0.422
53	0.179	0.228	0.271	0.320	0.352	0.418
54	0.177	0.226	0.268	0.317	0.349	0.414
55	0.175	0.224	0.266	0.314	0.346	0.411
56	0.174	0.222	0.264	0.311	0.343	0.407
57	0.172	0.220	0.261	0.308	0.340	0.404
58	0.171	0.218	0.259	0.306	0.337	0.400
59	0.169	0.216	0.257	0.303	0.334	0.397
60	0.168	0.214	0.255	0.301	0.331	0.394

11 Data CD

The attached CD contains, besides the orignal data, data files used for the calculations as well as those data ranges which the graphs and diagrams are based on.