Development of an Instrument and Evaluation Pattern for the Analysis of Chemistry Student Teachers' Diagnostic Competence

Yannik Tolsdorf
University of Bremen
y.tolsdorf@uni-bremen.de

Silvija Markic*
Ludwigsburg University of Education
markic@ph-ludwigsburg.de

Abstract
Diagnostics have become one of the central tools for planning teaching which accommodates learners' needs. Changes occurring in the chemistry classroom caused by rising levels of student heterogeneity play a crucial role. Thus, diagnostic processes are seen to be part of the cure. For this reason, chemistry (student) teachers need to understand the possibilities and potential for diagnostics in their classrooms. The present study employs a definition of diagnostic competence by Jäger. He emphasizes three domains: conditional knowledge, technological knowledge and knowledge of change. To achieve such competences, student teachers need to start learning about it during their university training. Since there are only few and mainly quantitative instruments in this field, the present paper describes the development of an instrument and an evaluation pattern for analyzing student teachers` diagnostic competence in chemistry. The instrument is a qualitative approach. It is based upon a written essay and open-ended questions. The evaluation pattern will be in focus. Initial results from the present study will also be discussed.

Keywords
Diagnostic competence, evaluation pattern development, competence development

◆Received 29 August 2017 ◆Revised 21 November 2017 ◆Accepted 1 December 2017

State of the literature

- The diagnosis and support of students is a good possibility to consider the heterogeneity of the class for chemistry teaching.
- The focus of the teachers´ diagnostic competence was often the (in)correct judgement of the teachers´ assessment of students, but the research also describes different results. On the one hand, a good (but still improved) correlation is described. On the other hand, the standardized tests don´t correlate with the judgement.
- The changes of the diagnostic knowledge in the teacher training program in chemistry hasn´t been evaluated.

Contribution of this paper to the literature

- The focus is on the investigation of the knowledge about the formative diagnosis and not on the final correct assessment.
- The development of a questionnaire which helps to investigate the attitudes, beliefs and skills of student teachers in chemistry education regarding to diagnosis, heterogeneity and diversity.
- In additional to the questionnaire and this development, the evaluation pattern is the second focus of this paper to evaluate the data of chemistry student teachers.

Introduction
When researching and discussing teachers` knowledge levels, the concepts of Pedagogical Content Knowledge (PCK), Content Knowledge (CK) and Pedagogical Knowledge (PK) are usually cited. The construct of Pedagogical Content Knowledge (PCK) is widely used in science education research. It was mentioned for the first time by Shulman (1986). PCK is specific professional knowledge, which is developed and expanded upon during teacher education and intensively reflected upon during work experience. Such knowledge includes both measuring beginning students' foreknowledge and explicitly understanding their individual characteristics as learners. This includes how to personally define and diagnose such factors among learners (Hashweh, 2005; Loughran, Berry & Mulhall, 2006; Shulman, 1987). Park and Oliver (2008) and Olszewski (2010) see student conceptions and misconceptions as an aspect of PCK. On the one hand, this knowledge helps to prognosticate student responses. On the other, it helps to diagnose such student conceptions. Learning conceptions and misconceptions are important elements in the planning and design of science teaching, in particular chemistry lessons (Barke, Hazaari & Yitbarek, 2009). In order to diagnose such concepts in scientific educational studies, a large number of diagnostic instruments have been developed over a long period of time (e.g. Barke, Hazaari & Yitbarek, 2009; Kahveci, 2013; Peterson, Treagust & Garnett, 1989; Taber, 2002).

During the last few years, diagnostics have become increasingly popular in both general and science education research (Chandrasegaran, Treagust, & Mocerino, 2007; Klug, 2011; Vogt & Rogalla, 2009; Wagner, Göllner, Helmke, Trautwein, & Lüdtke, 2013). However, even by the early 1980s diagnosis and diagnostic skills had already been explicitly mentioned in the research literature (Coladarci, 1986; Gillespie, 1991; Shulman, 1986). There exists a very high correlation between diagnostics and a high level of teaching and learning effectiveness (Fischer et al., 2014).

But diagnostics and diagnostic skills in science education are not only spoken about in the context of misconceptions. This is especially true for the ongoing discussion about diversity and inclusion in science classes. Diagnostics and diagnostic skills have been linked to concepts including: (i) handling heterogeneity (Grossenbacher, 2010), (ii) facilitating inclusion (Florian & Black-Hawkins, 2011), (iii) developing teaching units (Vogt & Rogalla, 2009), (iv) focusing on the individual support of students (Barke, Hazari & Yitbarek, 2009), (v) diagnosing learning disabilities (Williams, 2013), (vi) measuring teacher competence (Loughran, Berry & Mulhall, 2006) and (vii) creating linguistically sensitive science lessons (Markic, Broggy & Childs, 2012; Tolsdorf & Markic, 2016a). These different contexts of using the concept of diagnostics in school should show that diagnosis is a fundamental tool for all school subjects. But what do diagnostics mean for the school system overall and, consequently, for chemistry teacher training courses? Which knowledge about appropriate diagnosis in the classroom must chemistry teachers and teacher trainees possess in order to succeed? And how can we measure such diagnostic competence?

https://doi.org/10.12973/ijpce/80707
* Corresponding author

Theoretical Background

The definition of diagnostics in the school context as given by Ingenkamp and Lissmann (2008) contains a broad, yet in-depth understanding of diagnosis that can be used in many areas. The authors state that "diagnosis includes all diagnostic activities which evaluate the conditions and the skill set of the individual or group learner, which can be observed during a planned teaching and learning processes" (Ingenkamp & Lissmann, 2008, p. 13, translated from German by the authors). In a school context two directions are normally mentioned: 1) the pedagogical-psychological diagnostic and 2) the pedagogical-didactic diagnostic (Füchter, 2011). Of particular relevance for teaching a given subject is the pedagogical-didactic diagnostic. This area covers the diagnosis of learners' starting points, their specific learning processes and their student performance (Füchter, 2011).

In general, diagnosis should be seen as a cyclical process. Heidemeier (2005) describes the diagnostic process as cyclical, since the implementation must be reevaluated at the end of each cycle. Many subsequent cycles of diagnostics can and should be performed, since new questions may arise during the process. Klug, Bruder, Kelava, Spiel, and Schmitz (2013) present a three-step cyclic process in their model of the diagnostic process (see **Figure 1**). At the start, the diagnostician faces a problem or question which needs to be answered by the end of the process (Füchter, 2011). Thus, teachers should also define the aim of diagnosis at the beginning of their efforts. To fulfill this aim, special methods and instruments need to be selected (pre-actional phase). Following this comes the actional phase, in which data are collected. The post-actional phase represents the end of the first round, in which a promotional plan is developed and implemented. This three-step cyclical model helps teachers when planning and implementing diagnosis in their lessons.

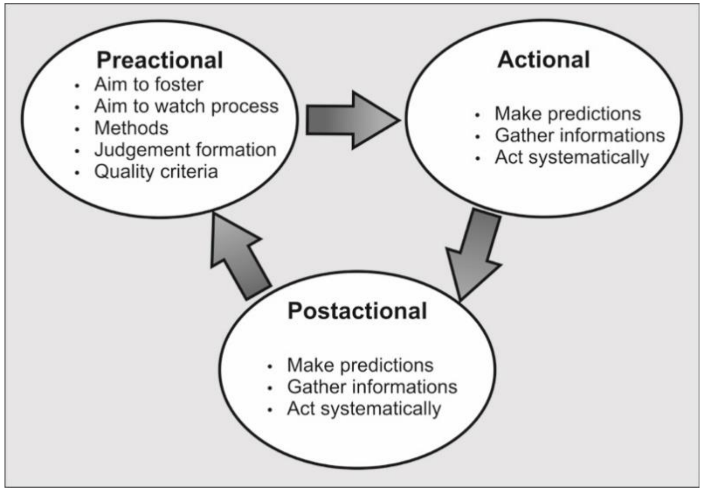

Figure 1. The diagnostic process by Klug et al. (2013)

Furthermore, the aim of the diagnosis must be clear at the beginning of a pedagogic-didactic diagnostic. Basically there can be two possible main focuses for this kind of diagnostic test: final assessment of learning processes and lessons-related performance diagnosis (Schrader, 2013; see **Figure 2**). The first type of diagnostic is designed to measure students` learning success and to grade them at the end of a teaching unit. Marks are used to assign authorizations (certificates) and the results of such diagnosis can have long-term consequences for students (Brookhart, 2011). The second focus involves the monitoring and optimizing of teaching and learning processes (Nitko & Brookhart, 2007).

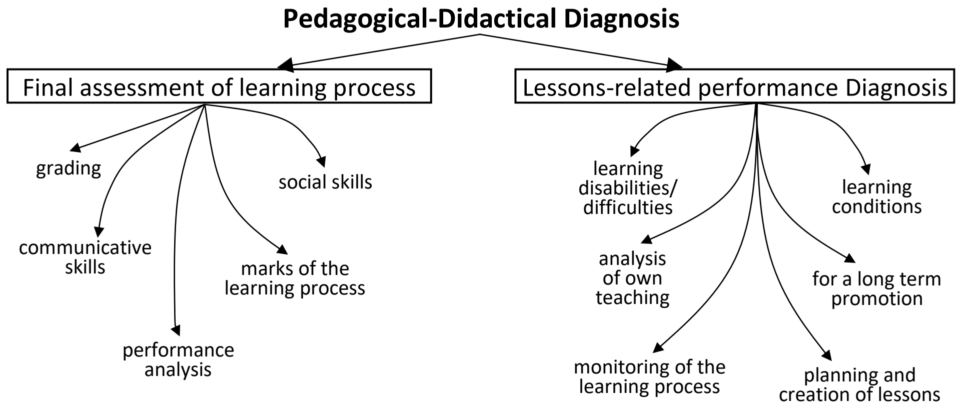

Figure 2. Structure of the pedagogical-didactic diagnosis (Tolsdorf & Markic, 2016a)

This very brief overview of diagnosis in the school context has shown that teachers should have a broad and deep understanding of it. Thus, in order to successfully carry out proper diagnostics in school, for example, analyzing the linguistic skills of pupils in heterogeneous classes, the teacher must possess the relevant diagnostic skills and qualifications. This is far more than the knowledge about adequate methods and instruments which can be used to answer a certain question.

Diagnostic competence is repeatedly discussed in the context of PCK (e.g. Tolsdorf & Markic, 2016b). PCK includes knowledge about students' foreknowledge and an explicit understanding of students` individual characteristics, including how to define and diagnose such factors (Loughran et al., 2006). Furthermore, four of the six categories from the differentiated hexagon model of PCK by Park and Oliver (2008) include aspects of pedagogical diagnosis. Krauss et al. (2004) defined diagnostic competence not just as one single competence, but rather as multiple facets of competencies. Consequently, diagnostic competence cannot merely be described as a sub-component of PCK. Schrader (2013) defines diagnostic competence and focuses on two main statements. First, diagnostic competence is the teacher's ability to successfully cope with the upcoming tasks. Second, Schrader (2013) focuses on the quality of the diagnosis, which tends to be the main focus when educators discuss this issue. This explains why diagnostic competence has mainly been studied in view of teachers' ability to accurately judge learners or to achieve a "correct" diagnosis. This has been the case since the 1970s (Coladarci, 1986; Feinberg & Shapiro, 2009; Klug, 2011; Perry, Hutchinson, & Thauberger, 2008). However, the terms "diagnostic competence" and "diagnostic skills" have long established themselves in the research literature (Krauss et al., 2004). In general, Jäger (2006) characterizes diagnostic competence with the help of six knowledge domains. The sixth knowledge domain (pychodiagnostic competence) is a global competence and does not play a role in the school context. The five important domains for teachers are:

1. Conditional knowledge – knowledge about the given background of one person and the influences that affect this person's specific experiences or cause certain behaviors. In addition this includes knowledge of such effects and their possible manifestations in a given survey.
2. Technological knowledge – the ability to select the most appropriate data collection techniques and to choose proper analysis methods for diagnostic questions.
3. Knowledge of change – knowledge development which includes the application of strategies dealing with changing learners' experiences and/or the behavior of those involved in the interactions.
4. Knowledge of the comparison – knowledge about the classification of behavior within a comparative group.
5. Competence knowledge – possession of sufficient knowledge to be able to answer a specific question. If the teacher does not possess this, then he or she must either expand his knowledge or seek assistance from a more competent person.

The first three dimensions are being the most important for teachers (Tolsdorf & Markic, 2016a).

Jäger (2006) views the last competence as important only for psychologists; however, Füchter (2011) disagrees on this point. Füchter states that this knowledge domain is important for school life, because it protects teachers from being overtaxed in their abilities. For example, teachers should not perform diagnosis of dyslexia unless they have been specially trained for it. Instead they should seek help. This frees up teacher time and resources, while still providing expert aid and a final answer to the question.

As already mentioned above, many studies have been limited in their understanding of teachers` diagnostic competence with respect to the teacher's accuracy of judgment. Often the correlation between a teacher's personal judgment and reality is analyzed using less distortion-prone test instruments (e.g. Demaray & Elliot, 1998; Feinberg & Shapiro, 2003; Hoge & Coladarci, 1989; Partenio & Taylor, 1985). It has already been shown that teachers can assess the ranking of student performance rather precisely, but that their individual judgment varies widely from what really happened. Bates and Nattelbeck (2001) were able to show that teachers tend to overestimate the reading performance of their students (especially struggling readers).

Starting from the present we can see the importance of diagnostic competence for teaching in general and chemistry in particular. In order to achieve Jäger's (2006) knowledge domains of diagnostic competence as defined above, development of such competencies should be begun during university teacher training programs. Thus, the development of diagnostic competence should play one of the crucial roles in both overall teacher education and in subject-specific training efforts. To approve the influence of university seminars and lecturers on the development of student teachers` diagnostic competence in chemistry, it is important to find an instrument to evaluate their competencies. However, very few studies to date have examined the development of diagnostic competence during teacher training. Even less have specifically looked at chemistry teachers. That is why our project created university modules to develop student teachers` diagnostic competence in chemistry. The first aim of this project was to develop an evaluation instrument. In following discussion below, the development of the instrument and the evaluation patterns with their initial results will be presented.

Instrument

In order to answer the research question, a questionnaire was developed. The student teachers were asked for relevant information such as their age, sex, the number of semesters studied, etc. Furthermore, information about their migration and linguistic backgrounds and personal knowledge of foreign languages was collected. This was important, because such information aids in understanding the participants' views on the identification, perception and handling of the heterogeneity and diversity within the chemistry classroom (see e.g. Moore, 2007).

Because theories of diagnostic competence in science education remain few and far between, the following research was based upon open-ended questions. Another reason for this exploratory study with the help open-ended questions is that students could express their attitudes, beliefs and opinions on the subject (this is not influenced by the "social desirability effect" (see Weisberg, 2005)). Thus, the second part of the questionnaire began with the task "Write an essay about diagnosis in chemistry lessons." The idea to start with an essay was not introduced to influence the participants on this issue in any way. Instead it was aimed at collecting a time-dependent snapshot of learners' first-hand knowledge and beliefs about this topic. By using such an open-ended, broad questioning method, students could be asked to give information on aspects that are not described in the literature.

To ensure that the pertinent research questions were answered, a third part of a questionnaire was added. This part contains four specific, open-ended questions which are based on the first three knowledge domains by Jäger. Thus, the focus remains on the three main aspects of diagnostic competence which are essential for chemistry teachers (see Tolsdorf & Markic, 2016a). The four research questions are phrased as follows:

1. Describe how the heterogeneity of a learning group can affect education.
2. Describe what methods you would use for diagnosis (e.g. the language level of the child).
3. Describe what strategies you would use in the classroom to deal with heterogeneity.
4. Describe how (or if) you would include heterogeneity in your lesson planning.

According to Jäger (2006) teachers need knowledge of possible factors which influence chemistry teaching (compare to the first question in the third part of the questionnaire). The main factor in question is the learning conditions faced by pupils. These are included in the terms heterogeneity and diversity. This question is limited to heterogeneity, because all student teachers understand this. Furthermore, both terms are used as synonyms in everyday life.

The second open-ended question relates to technological knowledge as defined by Jäger (2006). Thus, student teachers should know about diagnostic methods, diagnostic instruments and analytical methods. They should be able to describe the use of these methods for chemistry teaching and learning.

The last important domain is the knowledge of change. According to Jäger (2006), this knowledge domain consists of two parts. First, the students must possess knowledge of strategies for dealing with heterogeneity. Information about this aspect is requested in the third question in the questionnaire (see above). Moreover, student teachers need to understand changes in knowledge occurring in the classroom in order to use relevant strategies to teach and plan lessons. This is covered by the last question in the questionnaire.

Context of the study and Sample

To answer the research questions, data were collected from participants at different stages of their university teacher training program. The focus was on two teaching modules aiming at the development of diagnostic competence of university students: (i) Module Chemistry Education 2 (ChemEd 2) and (ii) Module Chemistry Education 4 (ChemEd 4) (see figure 3). The foci of the modules are based in three steps: (i) sensitization of chemistry student teachers for heterogeneity and diversity, (ii) diagnostic of students' individual requirements and need and (iii) dealing with heterogeneity and diversity in chemistry classes.

Bachelor students visit Module Chemistry Education 2 and attend their second official course in the didactics of chemistry. Before this time, they have not had any seminars explicitly dealing with diagnosis. They also have no previous school internship experience. The Chemistry Education 2 module consists of two seminars which cover 1) planning and diagnosing and 2) various methods in chemistry teaching. After this module, teacher trainees have an in-school internship which includes twelve teaching hours. Module Chemistry Education 4 is a part of their Master's degree. Student teachers attend a seminar about students` preconceptions, then attend a second internship lasting four months with fifteen hours a week of teaching. The purpose of this pilot study was a development of the instrument and an evaluation pattern. Thus, **Figure 3** presents the time points of data collection: (i) before ChemEd 2 and (ii) after ChemEd 2, but before ChemEd4. Both modules start at the same time. The modules ChemEd2 and ChemEd4 are described more in detail by Authors (2016b).

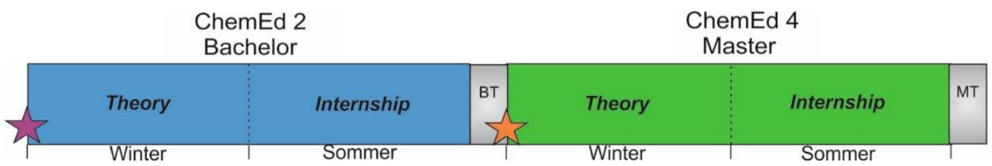

Figure 3. Time points of data collection (BT= Bachelor's thesis, MT= Master's thesis)

The second group of student teachers (orange star) had already completed ChemEd 2 and had just begun ChemEd 4. A total of twenty-eight student teachers completed the questionnaire. This group consisted of seven males and twenty-one females. Biology and mathematics were also the most common second-subject combinations with chemistry in this group. The student teachers were all older than twenty-one, but eighteen of them were still under 25 years old. Two student teachers had a Russian migration background.

The development of the evaluation pattern

Quality concerns play a central role in all stages of qualitative studies, beginning with the verbal expression of research questions, extending into data collection, and ending with the analysis and interpretation of research findings. For the present study the quality criteria for qualitative research were tested with the help of concepts developed by Mayring (2014) and Kippendorff (1980). First of all the "documentation of methods" approach was strictly adhered to. All work steps and decisions made during the collection, transcription and analysis were recorded, so that the research process remained understandable. In this paper only the most important steps will be described. The first step behind the development of a questionnaire was the development of an evaluation pattern following qualitative content analysis by Mayring (2014) using the evaluation program, MAXQDA. Taking Jäger's four knowledge domains of diagnostic competence as a lead, the categories and sub-categories could be developed inductively from the data. A coding unit was defined as one train of thoughts. After 50% of the questionnaires were coded, the developed evaluation pattern was revised with the help of additional people, so that the categories were clear and the level of the abstraction of the categories matched the research questions. The revised evaluation pattern allowed a new coding of all the data to be performed. After the data were coded, the evaluation pattern was checked for completeness. Each of the coding units could then be assigned to a code from this developed evaluation pattern. The individual codes were discussed and defined by small groups of researchers. An example case illustrates the coding technique and serves to help people inexperienced in such data analysis methods with their work. Therefore, the data could be analyzed with the help of other people using the developed evaluation pattern.

The development of this study was data-oriented, product-oriented and process-oriented. Semantic and sampling validity was reached by appropriate definition of categories and sub-categories. We started with the data and tested them with the aid of two experts. Furthermore, we could communicatively validate the results because we discussed each step of the development of the evaluation pattern with experts. We were also able to reach correlative validity (triangulation) for the present study thanks to the interviews with the student teachers. This part of the study is presented in Tolsdorf and Markic (in preparation).

Finally, two independent researchers performed the data analysis. The codes were compared in order to reach an inter-coder agreement as defined by Swanborn (1996). Our value turned out to be 87%, which is very good. The two coders dealt with rare instances of disagreement. Thus, stability and reproducibility could also be upheld. The created categories gave a good representation of the data, since the inter-subjective agreement for a qualitative research is high. These categories are described in more detail in the evaluation pattern section below.

The evaluation pattern

We developed an evaluation pattern inductively from the data. First, four domains were identified and defined for diagnostic competence following Jäger's (2006) ideas: (i) competence knowledge, (ii) conditional knowledge, (iii) technological knowledge and (iv) knowledge of change. Then step-by-step formulation of the inductive categories for the collected data was performed. These categories explain the domains in a more detailed fashion. Finally, sub-categories for each category were created from the data. Only the sub-categories which were mentioned by the majority of the student teachers were chosen as relevant for the present project. An example from the data was selected for each sub-category. These examples for the developed categories and definitions in the domain of "competence knowledge" are presented in **Table 1**. In this case, no sub-categories were found. Competence knowledge can be seen here as awareness of, attitudes towards, and beliefs about diagnostics at school. This includes the teacher's ability to improve his or her knowledge and seek support from experts.

Table 1. Definition of the categories of competence knowledge, explained an example (translated from German by the authors).

Category	Definition	Anchor Sample Statement
Insecurity in knowledge about diagnosis	Student teacher comments on the need of diagnosis in school, but he/she does not feel able to perform it.	*"I wish to have one tool for diagnosis which I can use easily in my lesson. I don't know any another way to discover the problems."*
Sensitivity for diagnostics	Student teacher expresses the need for diagnostics in chemistry lessons, gives reason(s) for the need but does not name any examples.	*"If you can diagnose your students' needs and problems (which is sometimes really difficult) you can use the results for planning your chemistry lessons."*
Reasons for diagnostics	Student teacher names single reasons for the process of diagnostics in chemistry lessons, but he/she does not describe a diagnostic process as such.	*"The results of the diagnosis can give you a hint where to start to help the student."*
Awareness of a diagnostic as a process	Student teacher describes both a diagnostic process and its meaning. He/she mentions the purpose of the diagnostic.	*"During the lesson the diagnostic has to take place. You need to check your lessons regularly. By doing so, you can change your lesson module to suit student needs."*
A wish for more skill/ knowledge	Student teacher mentions the need for more skills or competencies for diagnosing in chemistry classes.	*"Heterogeneity changes chemistry lessons. More qualified teachers are needed …. including myself."*
Attitude towards heterogeneity	Student teacher expresses his/her feelings towards heterogeneity in chemistry classes. These can be positive, negative or mixed. The attitude towards heterogeneity can be established on the basis of consequences, which are formulated in connection with the conditional knowledge.	*"If students cannot work properly, because some people are faster and some slower, they get frustrated. However, the differences can also be seen as an advantage. Students can also help each other."*
Importance of diagnosis	Student teacher specifically expresses the importance of diagnostics or the diagnosis process for chemistry teaching.	*"Diagnostic is one of the main points in being a teacher. It is a very important topic for teachers."*

The second domain is conditional knowledge. This describes factors that cause students to have different experiences and which influences their behavior in chemistry classes. This domain therefore includes the relevant knowledge factors affecting pedagogical conditions. Three main factors of influence are given inductively by the data. The three categories and their explanations are as follows:

i. "Individual influences of or by pupils" means the specific characteristics of individual students, which can affect teaching and the learner's behavior in the chemistry classroom. In this context, student teachers mention different dimensions of the diversity wheel (Diversity Leadership Council of the John Hopkins University) which can influence chemistry teaching and learning. Single dimensions are named according to their influence, but are not explained in detail.
ii. "Administrative and organizational influences" include external aspects which are imposed by the country or school system which affect teaching and learning. Student teachers name single factors that are not part of the lesson, but rather imposed upon them by either the administration and/of political system. Single reasons are named, but their influence is not explained in detail.
iii. "Influence by the lesson and its planning" describes the factors which affect students' learning and their motivation. They are caused by the lesson itself and the overall planning of the lesson. Student teachers name specific factors from a lesson which can influence student behavior.

Based on the named categories and on the data, sub-categories were developed for this domain. The categories and sub-categories for conditional knowledge are presented in **Table 2**.

Table 2. Sub-Categories for conditional knowledge, including examples (translated from German by the authors).

Sub-Category	Anchor Sample Statement
Individual influences of students	
Linguistic heterogeneity	*"Linguistic heterogeneity can cause problems in understanding."*
Migration/Immigration	*"Migration backgrounds can influence learning behavior."*
Learning difficulties	*"Children with special needs (dyslexia or emotional and behavioral disorders) can complicate the understanding of materials or statements in the classroom, respectively. This can make learning difficult or impossible."*
Educational biographic/socio-economic background	*"Topics are regarded through the lens of different social, linguistic and particularly educational biographical backgrounds."*
Cultural diversity	*"Each student brings their own culture and their own beliefs and traditions to the class. This should be considered in the planning of a lesson."*
Heterogeneity of learning effort/content knowledge	*"Heterogeneity in the level of learning effort can lead to the demotivation of students for whom the teaching tempo is to slow or fast. They are either overwhelmed or bored by the pace."*
Physical disability	*"Physical disability might need to be considered in the lessons (for example during experiments)."*
Motivation	*"The aspect of general learning behavior, including interest in the subject, matter or object of interest, needs to be diagnosed."*

Table 2. (*Continued.*)

Administrative and organizational influences	
Lack of (effective) lesson time	*"Difficulty for teachers arises because they often teach a subject two hours per week, which are widely scattered temporally."*
Number of students in the classroom	*"Classes should not be that overcrowded, if accurate diagnosis is to occur."*
Influence by the lesson and its planning	
Social behavior	"Strife between students can occur and disrupt learning."
suboptimal support	*"The "clever (students)" are under-challenged; the "non-clever (students)" overwhelmed. It is not easy to find the golden mean."*
Changes of the lesson plan	*"The teaching did not follow the advance lesson plan. Heterogeneity has a big influence on the lesson in my opinion."*
multiple and different ideas	*"There are many different opinions in one topic which are caused by heterogeneity. This is good. The discussion on the topic is more intense."*

The third domain is called technological knowledge. This knowledge is needed for the preparation and the implementation of diagnostics (the pre-actional and actional phases). For this purpose, this domain describes knowledge about data collection possibilities and the analysis of data. However, student teachers only mentioned different instrument for data collection, not methods of analysis. Therefore, we only coded for the category "Methods of data collection". All of the different methods mentioned were listed in sub-categories. These sub-categories are presented in **Table 3** with examples.

Table 3. Sub-Categories of evaluation methods from technological knowledge (translated from German by the authors).

Sub-category	Anchor Sample Statement
Games	*"Crosswords getting information from the game."*
Intuitive action	*"I guess I probably did it rather intuitively."*
Observing	*"Observing the lesson exactly and, if possible, making statistics."* *"Observation is an ideal possibility to get more information about your own lesson."*
Communicating with and between students	*"Communication with your own students"* *"Observing the communication between students"*
Presenting	*"Reporting on and repeating the learning topic by presenting, reporting in front of the class, seminar papers, … (dealing with scientific language)"*
Reflecting	*"Self-reflection or reflection in an interview"*
Writing (any kind of worksheets)	*"Students must write a text, for example protocols in chemistry education"*
Testing	*"Quizzes"* *"Look for language difficulties in written exams."* *"Class test."*

The last domain is the knowledge of change. It describes the methods and ideas for assessing changes in the chemistry classroom while dealing with heterogeneity of students. Four categories with were developed from the data:

i. "Changes during the planning of the lesson" describes the possibilities or aspects which can be considered while planning your own chemistry lesson.

ii. "Changes of teacher behavior" contains all methods that could change teacher's actions.

iii. "Changes of the teaching materials" names different possibilities for the adaptation of teaching and learning materials to students` needs.

"Changes of the framework" does not relate to teacher activity, but rather to administrative changes. These sub-categories are shown in **Table 4** with examples.

Table 4. Sub-categories from the knowledge of change (translated from German by the authors).

Sub-category	Definition	Anchor Sample Statement
Changes during the planning of the lesson		
Choosing topics and aims for students` need	Student teacher describes teaching as based on the respective students. Therefore it must include the opinions, interests and motivations of the students.	*"If you know about your students, then you can try to take topics that have relevance for the most pupils; or you can offer to choose more themes/topics."*
Planning of effective learning time	Student teacher mentions effective teaching time, which should be reconsidered after the diagnosis (data collection). It is stated that students learn differently and need different amounts of time.	*"For slower students, regularly incorporate repetitions in your planning, since they need more time for learning. In the same class, prepare reserves and few more topics for faster students. If you don't know how fast the class is, plan different exits out of the lesson."*
Teaching methods	Student teacher names different teaching and learning methods and social forms of teaching.	*"Use of cooperative methods."*
Forming groups	Student teacher describes different group structures in the class, so that the pupils can learn in the best possible way.	*"Form learning groups that are mixes of faster and slower learners. High and low achievers should be in the same group.*
Differentiation	Student teacher lists various methods for differentiation in chemistry teaching (for example differentiation, individualized lessons).	*"Designing individual learning plans."* *"Apply internal differentiation for reaching all of the students."*
Different ways of learning	Student teacher considers the different learning styles of the students in the lesson plan.	*"This means that different types of learning to be included in the planning of the lesson."*

Table 4. (*Continued*)

Changes of teacher behavior		
Teacher language	Student teacher mentions that the teacher should modify their own language (written and spoken) when explaining or in writing on the blackboard and modify their language in worksheets.	*"The teacher must use understandable language so all of the students understand the explanations."*
Help from colleagues	Student teacher describes teachers seeking more support from colleagues or other people in school (social workers).	*"Likewise, it is important to have your teaching examined by other colleagues/people."*
Changes of the teaching materials		
Help for learning a content	Student teacher lists teaching and learning aids (for example, help cards).	*"Exercises with different levels of challenge and various ways to the object of study are offered to a heterogeneous class."*
Help for linguistic skills	Student teacher mentions linguistic aids as teaching support related to language.	*"Assistance at the linguistic level."*
Design of materials	Student teachers describe the development of teaching and learning materials according to the needs of the students.	*"The design of materials must be adapted to the needs of the students."*
Design of experiments	Student teacher describes adapting experimental instructions to the needs of the students.	*"Differentiated designed experiments, for example, those supported by pictures."*
Additional materials	Student teacher mentions that teacher should create additional materials that are not designed for the normal lesson.	*"More materials for fast pupils."*
Changes of the teaching Framework		
Considering needs of disabled students	Student teachers states that the school, classroom or laboratory needs to be barrier-free, so that all students (e.g. students in a wheelchair) can follow and experience the chemistry lesson.	*"The concept of teaching must be adapted to physically handicapped students, e.g. less movement during the lesson or thinking about the amount of experiments."*
Parents	Student teacher says that parents should be involved in the diagnostic process as an adviser and helper.	*"For example, discussion with parents."*

Results

The results of the study are presented below. However, the focus remains only on the development of the questionnaire and the evaluation pattern.

For competence knowledge we can identify differences between the two groups and the time points of data collection. Almost half of the student teachers had more negative attitudes toward heterogeneity in chemistry classes before they attended any of the learning modules. The group pointed out especially the organization and the conduction of the laboratory sessions. They assume that - depending on students` heterogeneity and the experiments – laboratory work could be very difficult and sometimes impossible. They viewed it as a problem that needed to be dealt with. However, the second group didn't mention this issue at all. With respect to awareness of the diagnostic process in chemistry classes, the student teachers in the very first group mentioned the need for diagnosis much more often than the more advanced student teachers. It seems that increasing levels of competence knowledge affect a change in teacher attitudes toward heterogeneity and diversity. Delving deeper into the data, we can see which aspects of heterogeneity and diversity are mentioned by the student teachers most often. Both groups in this study mentioned students` linguistic skills and differences in students` content knowledge and their misconceptions as important influences on heterogeneity in the classroom. These are both important conditions for learning chemistry. Other dimensions of the diversity wheel appeared to be unimportant to student teachers in the beginning phases, but become more important after they took part in their first internship. The student teachers in the second group mentioned these dimensions as well. Interestingly, student teachers thought that "no appropriate support of students" was a reason for diversity in chemistry lessons after their first practical phase.

We can see that the highest levels of coding were in the sub-categories for technological knowledge, but the focus remained in only two or three sub-categories for all of the student teachers. All of the student teachers in this study focused primarily on written data collection. Written means any kind of worksheets with different topics or even experiments. One student teacher wrote (translated from Germany by authors): "Protocols/Lab reports of students can be collected to diagnose language" or "Speech balloons (or thought bubbles) about reactions (or reaction arrow) can help to diagnose learners". The first group of student teachers mostly mentioned that tests as a diagnostic strategy seemed unimportant to them. This group focused more on classroom observation and interviews with or between students. The second group, however, mentioned that tests are an important strategy for diagnosis. In contrast, observation seems to be viewed as a relatively unimportant tool for diagnosing students during the course of the study by the members of this group. At the beginning of ChemEd 4, interviews showed lower coding levels than in the other time period. These student teachers mentioned games or play as a tool for diagnosing students in the chemistry classroom after their internship.

The knowledge of change was the most prominent aspect in both groups of student teachers. The predominate strategy in all of the groups was differentiation; however, this seemed to markedly increase by the end of the first internship. Furthermore, data showed that student teachers saw effective teaching and learning time as factors that need to be taken more seriously into consideration. The same can be said for the importance of matching worksheets to learners` needs and about diversity within the class itself. Another aspect in this knowledge domain was that experiments, the layout of experiments and linguistic help cards were more important tool during practical study. Help cards for learning content or subject knowledge were, however, important for both groups. Student teachers only tended see talking with colleagues as a way to make a change in the chemistry teaching after they had completed their practical phases in schools.

Discussion and Conclusion

Chemistry student teachers in this study appeared to be sensitive to heterogeneity in the classroom and its influence on the learning atmosphere during lessons. On the one hand, they realized that heterogeneity can negatively influence the classroom atmosphere and cause disruptions. On the other hand, the same heterogeneity can aid in developing learners' social skills and in contributing to a better understanding of different cultures. It is good to see that such negative beliefs seem to change with experience. By second evaluation less student teachers defined heterogeneity or diversity as a problem and challenge. Generally, we observed that all chemistry trainees were sensitized to heterogeneity. We could also see that student teachers mentioned student's linguistic skills and content knowledge as main influences on heterogeneity in chemistry class. Other aspects on the diversity wheel also achieved increased relevance for the students during the course of their studies. Student teachers are focused mainly on written data collection at all stages of their teacher training program. Quizzes seemed to be important as well and take on added importance for older students. Conversely, observation appeared to take on later importance. The knowledge of change was the most pronounced aspect researched here. One reason for the higher levels of coding was the university seminar, because the modules very often focus on changes and developments in teaching, lesson plans and learner support.

The here presented instrument and the developed evaluation pattern seems to give a good overview over chemistry student teachers` diagnostic competence. It has been use in a meaning of cross-level and longitudinal study as well (Tolsdorf & Markic, 2017 and in preparation). The evaluation pattern was developed to mirror the data with the help of the inductive method. Thus, attitudes and skills, which are not fully detailed or described in the literature, could be collected from the participants. But open-ended questionnaires can also make student teachers view certain aspects (codings) of the topic as irrelevant to the question. This means that the participants may not mention something, even if they have knowledge about it. Jäger (2006) describes technological knowledge as the knowledge to select the most appropriate analysis methods for a data collection. The student teachers in the present study wrote nothing about analysis methods. Therefore, this knowledge domain could not be encoded in the evaluation pattern. From this starting point, future research might be well advised to explicitly ask about this point or to evaluating the university seminars in order to identify if this topic is present there or not.

Another critical issue is the knowledge of change, because this dimension by Jäger (2006) was included in two questions in the questionnaire. This dimension includes two aspects of educational action in the chemistry classroom. First, student teachers need knowledge about different teaching strategies. Secondly, they require knowledge about the use of strategies on the meta-level, so that they can reflect upon their own teaching actions. There were good reasons why this knowledge was divided into two separate questions. But the evaluation and interpretation of the data may have been influenced by this separation. This might explain the high coding levels for the knowledge of change in all the groups. One other explanation for the higher coding levels is probably seminar content. Many opportunities are discussed in each chemistry education module for how to plan and change lessons. The seminars and individual modules specifically focus on planning and changing chemistry teaching. This also explains a difference between the both groups in the knowledge of change. The students responded very differently to the two questions. Participants in the first group seldom saw a difference between these two questions and answered them in a similar manner. But the second round of data collection showed that the students highlighted different aspects of this dimension and were able to distinguish between the two questions. For these reasons, the two questions should remain separated in the questionnaire.

Another critical issue in the evaluation and interpretation of the data was that the "knowledge of comparison" was not evaluated for the group of chemistry student teachers in this study. Indications of this area were found in very few student answers in the first open-ended question. This confirms the assumption by Füchter that this knowledge domain is unimportant for teachers. Füchter (2011) (see also Tolsdorf & Markic, 2016a) says the three diagnostic knowledge domains (Conditional Knowledge, Technological Knowledge, Knowledge of Change) are particularly important for pedagogical-didactic diagnostic. Any differences between the dimensions by Jäger (2006) and these results (or Füchter, 2011) are due to the field of research from the researchers. Pedagogical-didactic diagnostic differs from educational-psychological diagnosis. This leads to a change in the level of importance of the knowledge domains for the diagnostician.

Finally, we can say that this instrument isn't perfect in the sense of a general definition of diagnostics by Jäger, yet many aspects could be collected from the student teachers. This was especially true for the focus on the pedagogical-didactic diagnostic. Further studies with and extensions of the instrument are needed. However, for the purpose of the current study and in view on the research questions, we can say that the instrument and the developed evaluation pattern are satisfactory. The development and differences in the diagnostic competence of chemistry student teachers could be examined by this pilot cross-level case study. Through this development of an instrument, gaps and problems in diagnostic competence were discovered in the teaching modules. This means that the modules need to be developed further. Overall, this questionnaire proves to be a good tool, because it complies with the quality-control criteria for good qualitative research (Mayring, 2014) and delivers good inter-rater value. At last it has to be said, that the instrument does not give a picture about student teachers` behavior and dealing in the real situation. Some more studies in this direction are needed in the future.

Acknowledgement

We thank the Deutsche Telekom Stiftung (German Telekom Foundation) for the financial support of this PhD-project.

References

Barke, H.-D., Hazaari, A. & Yitbarek, S. (2009). *Misconceptions in chemistry: addressing perceptions in chemical education*. Berlin: Springer.

Bates, C. & Nettelbeck, T. (2001). Primary school teachers` judgments of reading achievement. *Educational Psychology*, *21*(2), 177–187, doi:10.1080/01443410020043878.

Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3–12, doi:10.1111/j.1745-3992.2010.00195.x.

Chandrasegaran, A.L., Treagust, D.F. & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, *8*(3), 293–307, doi:10.1039/B7RP90006F.

Coladarci, T. (1986). Accuracy of teacher judgment of students' responses to standardized test items. *Journal of Educational Psychology*, *78*(2), 141–146, doi:10.1037/0022-0663.78.2.141.

Demaray, M.K. & Elliott, S.N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*(1), 8–24, doi:10.1037/h0088969.

Diversity Leadership Council of the John Hopkins University (Eds.)(without year). Diversity wheel. Retrieved August 24, 2017 from http://web.jhu.edu/dlc/resources/diversity_wheel .

Feinberg, A.B. & Shapiro, E.S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, *18*(1), 52–65, doi:10.1521/scpq.18.1.52.20876.

Feinberg, A.B. & Shapiro, E.S. (2009). Teacher accuracy: an examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, *102*(6), 453–462, doi:10.3200/JOER.102.6.

Fischer, A., Hößle, C., Jahnke-Klein, S., Kiper, H., Komorek, M., Michaelis, J., Niesel, V. & Sjuts, J. (2014). *Diagnostik für lernwirksamen Unterricht* [Diagnostic for successfull learning]. Hohengehren: Baltmannsweiler.

Florian, L. & Black-Hawkins, K. (2011). Exploring inclusive pedagogy. *British Educational Research Journal*, *37*(5), 813–828, doi:10.1080/01411926.2010.501096.

Füchter, A. (2011). Pädagogische und didaktische Diagnostik: Eine schulische Entwicklungsaufgabe mit hohem Professionalitätsanspruch [Pedagogical and educational diagnostic: a task of school development with a professionalism]. In A. Füchter & K. Moegling (Eds.), *Diagnostik und Förderung* [Diagnostic and support] (pp. 45–83). Kassel: PROLOG.

Gillespie, R. (1991). String teachers' diagnostics skills and their students' performance competencies. *Journal of Research in Music Education*, *39*(4), 282–289, doi:10.2307/3345747.

Grossenbacher, S. (2010). Kompetenz und Professionalität entwickeln [Development of competence and professionalism]. In A. Buholzer & A. Kummer Wyss (Eds.), *Alle gleich – alle unterschiedlich! Zum Umgang mit Heterogenität in Schule und Unterricht* [all the same – all different! Dealing with heterogeneity in school and lesson] (pp. 162–168). Seelze: Kallmeyer.

Hashweh, M. (2005). Teacher pedagogical constructions: a reconfiguration of pedagogical content knowledge. *Teachers and Teaching*, *11*(3), 273–292, doi:10.1080/13450600500105502.

Heidemeier, H. (2005). *Self and supervisor ratings of job-performance: Meta-analyses and a process model of rater convergence*. Dissertation Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg.

Hoge, R.D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology*, *59*(3), 297–313, doi:10.3102/00346543059003297.

Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* [Textbook of pedagogical diagnostic], 6th Edition. Weinheim: Beltz.

Jäger, R.S. (2006). Diagnostischer Prozess [Diagnostic process]. In F. Petermann & M. Eid (Eds.), *Handbuch der psychologischen Diagnostik* [Handbook of psychological diagnostik] (pp. 89–96). Göttingen: Hogrefe.

Kahveci, A. (2013). Diagnostic assessment of student understanding of the particulate nature of matter: decades of research. In G. Tsaparlis & H. Sevian (Eds.), *Concepts of Matter in Science Education: Innovations in Science Education and Technology* (pp. 249–278). Dordrecht: Springer.

Krippendorff, K. (1980). *Content analysis. An Introduction to its Methodology*. Beverly Hills: Sage.

Klug, J. (2011). *Modelling and training a new concept of teachers' diagnostic competence*. Dissertation Thesis, Technische Universität Darmstadt.

Klug, J., Bruder, S., Kelava, A., Spiel, C. & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education*, *30*, 38–46, doi:10.1016/j.tate.2012.10.004.

Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W. & Neubrand, M. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz [COACTIC: professional knowledge of teachers, cognitively activating mathematics teaching and the development of mathematical competence]. In J. Doll & M. Prenzel (Eds.), *Die Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* [Educational quality of schools: teacher professionalisation, teaching development and student support as strategies of quality improvement] (pp. 31–53). Münster: Waxmann.

Loughran, J., Berry, A. & Mulhall, P. (2006). *Understanding und developing science teachers' pedagogical content knowledge*. Rotterdam: Sense.

Markic, S., Broggy, J. & Childs, P. (2012). How do deal with linguistic issues in chemistry classes. In I. Eilks & A. Hofstein (Eds.), *Teaching Chemistry – A Studybook* (pp. 131–156). Rotterdam: Sense.

Mayring, P. (2014). *Qualitative content analysis. Theoretical foundation, basis procedures and software solutions*. Klagenfurt: Beltz.

Moore, F. M. (2007). Language in science education as a gatekeeper to learning, teaching and professional development. *Journal of Research in Science Teaching, 18*, 319-343.

Nitko, A.J. & Brookhart, S.M. (2007). *Educational assessment of students*, 5th Edition. Upper Saddle River: Pearson Merrill Prentice Hall.

Olszewski, J. (2010). *The impact of physics teachers' pedagogical content knowledge on teacher action and student outcomes*. Berlin: Logos.

Park, S. & Oliver, J.S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284, doi:10.1007/s11165-007-9049-6.

Partenio, I. & Taylor, R.L. (1985). The relationship of teacher ratings and IQ: A question of bias?. *School Psychology Review, 14*(1), 79–83.

Perry, N.E., Hutchinson, L. & Thauberger, C. (2008). Talking about teaching self-regulated learning: Scaffolding student teachers` development and use of practices that promote self-regulated learning. *International Journal of Educational Research*, *47*(2), 97–108, doi:10.1016/j.ijer.2007.11.010.

Peterson, R.F., Treagust, D.F. & Garnett, P.J. (1989). Development and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching*, *26*(4), 301–314, doi:10.1002/tea.3660260404.

Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen [Diagnostic competence of teachers]. *Beitrag zur Lehrerinnen- und Lehrerbildung*, *31*(2), 154–165.

Shulman, L.S. (1986). Those who understand: knowledge growth in teaching. *Education Researcher*, *15*(2), 4–14.

Shulman, L.S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22, doi:10.17763/haer.57.1.j463w79r56455411.

Swanborn, P.G. (1996). A common base for quality control criteria in quantitative and qualitative research. *Quality and Quantity*, *30*(1), 19–35, doi:10.1007/BF00139833.

Taber, K.S. (2002). *Chemical misconceptions – prevention, diagnosis and cure: Volume I: Theoretical background*. London: RSC.

Tolsdorf, Y. & Markic, S. (2016a). Dealing language in science classroom – Diagnosing student` linguistic skills. In S. Markic & S. Abels (Eds.), *Science Education towards Inclusion* (pp. 23–42). New York: Nova.

Tolsdorf, Y. & Markic, S. (2016b). Exploring student teachers` knowledge concerning diagnostics in science lessons. In J. Lavonen, K. Juuti, J, Lampiselkä, A, Uitto & K. Hahl (Eds.), *Electronic Proceedings of the ESERA 2015 Conference. Science education research: Engaging learners for a sustainable future* (pp. 2002–2009). Helsinki: University of Helsinki.

Tolsdorf, Y. & Markic, S., (2017). Exploring chemistry student teachers´ diagnostic competence - a qualitative cross-level study. *Education Sciences*, 7(4), 86-100. doi:10.3390/educsci7040086

Tolsdorf Y. & Markic, S. (in preparation). Exploring student teachers´ competence concerning diagnostic in chemistry teaching – a longitudinal interview study.

Vogt, F. & Rogalla, M. (2009). Developing adaptive teaching compentency through coaching. *Teaching and Teacher Education*, *25*(8), 1051–1060, doi:10.1016/j.tate.2009.04.002.

Wagner, W., Göllner, R., Helmke, A., Trautwein, U. & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: dimensionality and generalizability of domain-independent assessment. *Learning and Instruction*, *28*, 1–11, doi:10.1016/j.learninstruc.2013.03.003.

Weisberg, H.E. (2005). *The total survey error approach: A guide to the new science of survey research*. Chicago: The University of Chicago Press.

Williams, A. (2013). A teacher's perspective of dyscalculia: Who counts? An interdisciplinary overview. *Australian Journal of Learning Difficulties*, *18*(1), 1–16, doi:10.1080/19404158.2012.727840.