International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Methodical challenges concerning the Draw-A-Scientist Test: a critical view about the assessment and evaluation of learners' conceptions of scientists

Bianca Reinisch , Moritz Krell, Susann Hergert, Sarah Gogolin & Dirk Krüger

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2017
https://doi.org/10.1080/09500693.2017.1362712

Methodical challenges concerning the Draw-A-Scientist Test: a critical view about the assessment and evaluation of learners' conceptions of scientists

Bianca Reinisch, Moritz Krell, Susann Hergert, Sarah Gogolin and Dirk Krüger

Biology Education, Freie Universität Berlin, Berlin, Germany

ABSTRACT
Students' and pre-service teachers' conceptions of scientists have been assessed in a variety of studies. One of the most commonly used instruments is the Draw-A-Scientist Test (DAST) which offers the advantage that no verbal skills are needed by the participants. In some studies, methodical challenges related to the DAST have been discussed; for example, the lack of drawing abilities among the subjects or the impact of the prompt itself on the drawings which lead to invalid interpretations. This study aims to evaluate the eligibility of the DAST or similar instruments to validly assess conceptions of scientists and their work. Pre-service science teachers in two cohorts ($N_1 = 79$, $N_2 = 101$) were prompted to draw their conceptions of a scientist, the location at which the scientist is working, and his/her scientific activity. The participants of cohort 2 were also asked to give written descriptions of their drawings and information on the conditions while they were drawing. From the results, several points of methodical criticism can be made which question the valid interpretation of DAST-drawings. Instead of using the DAST, we suggest developing and evaluating other formats, such as open-ended or closed-ended instruments which could produce a more validly interpretable assessment of individuals' conceptions in this field.

ARTICLE HISTORY
Received 3 March 2017
Accepted 29 July 2017

KEYWORDS
DAST; drawings; conceptions; nature of science; scientists; pre-service teachers; scientific activities; drawing assessment

Introduction

About 20 years ago, Finson, Beaver, and Cramond (1995) stated:

> If one of the goals of science education is to prepare students who are scientifically literate and to encourage students to pursue postsecondary study and careers in science and related fields, then students need to possess more positive images of scientists – the role in which they themselves may function in the future. (p. 201)

This statement is still deemed valid today. However, it should not be about positive, but about realistic images of science and scientists (e.g. Cakmakci et al., 2011). Wang (2013) conducted a study that reveals the importance of high school preparation in science and math courses to raise students' interest in choosing STEM-related post-school qualifications later on. The author suggests introducing and exposing students to

CONTACT Bianca Reinisch bianca.reinisch@fu-berlin.de Biology Education, Freie Universität Berlin, Schwendenerstr. 1, Berlin 14195, Germany

math and science courses in an early stage of high school education. Consequently, science teachers, among others, play an important role in shaping students' conceptions and it follows that a greater emphasis must be made on realistic images of science and scientists in science teacher education (McCarthy, 2015).

Since students' pre-conceptions are seen as a central component of planning lessons and teaching in class, they ought to be integrated here (Campbell, Schwarz, & Windschitl, 2016; Duit, Gropengießer, Kattmann, Komorek, & Parchmann, 2012). Studies show that teachers, similar to students, often hold conceptions of nature of science (NOS) aspects which are not in agreement with scientific perspectives (cf. Treagust & Duit, 2008) although an adequate NOS understanding is considered to be important for teaching students (cf. Lederman & Lederman, 2014; see below). Diverse influences shape the popular image and, consequently, learners' pre-conceptions of the scientific working field (cf. Finson & Pederson, 2011). One of them is media influences, such as fictional films which include a variety of different characterisations of scientists and their work. Popular conceptions of scientists and their working field may change over time due to the altering contemporary contexts of movie and show productions (cf. Pansegrau, 2008). Consequently, assessing prevailing conceptions should be a permanent research field in order to react to them adequately within science education.

One commonly used instrument to assess students' and pre-service teachers' conceptions of scientists is the Draw-A-Scientist Test (DAST) developed by Chambers (1983), in which test persons are prompted to draw a scientist. Finson et al. (1995) developed a corresponding checklist (DAST-C) to facilitate the analysis of the drawings. This form of assessment is frequently used in primary education research. Van der Veen (2012) suggested 'that the use of drawing for understanding is entirely appropriate for introductory college students' (p. 365). In a rising number of studies, it is now pre-service teachers' drawings that are being analysed to show their conceptions of science and scientists (e.g. McCarthy, 2015; Milford & Tippett, 2013; Subramaniam, Esprívalo Harrell, & Wojnowski, 2013). Other researchers (e.g. Losh, Wilke, & Pop, 2008) see the use of drawings in this context more critically, especially in regard to the validity of the interpretation of the drawings. The present article picks up on this point by introducing the DAST and by compiling advantages and disadvantages of this form of assessment. Methodical challenges will be analysed and discussed on this basis.

Theoretical background

Relevance of NOS to pre-service science teachers

With respect to Shulman (1986, 1987), three primary key components of professional teacher knowledge can be differentiated, which also have been established in education and educational research: pedagogical knowledge, content knowledge, and pedagogical content knowledge (cf. Grossman, 1990; van Dijk, 2014). An adequate understanding of NOS can be seen as part of teachers' content knowledge, whereas knowledge about approaches to teach NOS content in science classes can be associated with pedagogical content knowledge (Burton, 2013; Günther, Fleige, Upmeier zu Belzen, & Krüger, 2017).[1] For the development of the latter, a solid knowledge of the former is inevitable (Lederman, 2007; Lederman & Lederman, 2014; Reinisch & Krüger, 2016). Additionally, there are strong indications that teachers' attitudes toward science as well as their images about science and scientists directly influence their students' views (Christidou, 2011; Türkmen, 2008). It follows that NOS contents should be part of science teacher training in order for teachers to provide their students with an appropriate understanding of NOS. The latter is an objective that is highlighted in many standards documents in different countries (e.g. Germany: KMK, 2005; U.S.A.: NGSS Lead States, 2013) and, therefore, becomes ultimately an aim within science teacher education (Capps & Crawford, 2013; Krell, Koska, Penning, & Krüger, 2015).

'Draw a scientist!' – historical development

Ever since Mead and Métraux (1957) analysed 35,000 high school students' essays about their conceptions of a scientist, the so-called stereotypical view of a scientist has been a person who is male, elderly or middle-aged, wearing a lab coat and glasses, on his own in a laboratory performing experiments. In the subsequent more than 20 years, different studies applying different instruments have revealed the consistency of this stereotypical view of scientists, also across cultures (cf. Finson, 2002). A change in the assessment of students' conceptions was made by Chambers (1983) who developed the DAST, which is an instrument asking test persons to draw their conception of a scientist on a sheet of paper. Advantageously, researchers do not rely on the test persons' writing skills; a benefit of special importance when assessing younger children's conceptions (Losh et al., 2008). For analysing drawings, Chambers (1983) proposed seven indicators for the 'standard image' of a scientist: lab coat, eyeglasses, facial hair (beards, moustaches, abnormally long sideburns), symbols of research (scientific instruments, laboratory equipment), symbols of knowledge (books, filing cabinets), technology, and relevant captions (formulas, taxonomic classification, etc.). Chambers (1983) reported that at least one child in each class drew an alternative, mythic image (e.g. indications of danger, the presence of light bulbs, elements of mythic stereotypes such as Frankenstein creatures). He pointed out that alternative conceptions might be present in a larger proportion than assessed in his study. An additional prompt ('draw another scientist') in a small subsample of 24 test persons led to 11 students drawing the mythic scientist or elements of it.

Finson et al. (1995) developed the DAST-C on the basis of Mead and Métraux (1957) and Chambers (1983) work as well as on results of their own study. By using the checklist, Finson et al. (1995) aimed to improve the reliability in the analysis of drawings. The authors highlighted that the 'assessment is easy to administer and score, and no special materials need to be purchased, teachers can use the DAST-C in their classes to determine what, if any, stereotypical images their students hold' (p. 201). The checklist includes stereotypical elements which were detected in many previous studies. The more of these elements are to be detected in a drawing, the more stereotypical the conception of a scientist. The authors concluded that their checklist is a 'reliable, efficient, and effective format' (p. 204).

In the following decades, numerous studies were carried out addressing the conceptions of scientists relating to cultural background, age or socio-economic status, and the influence of different sorts of science instruction or out-of-classroom experience (for a literature review cf. McCarthy, 2015; Milford & Tippett, 2013). So far, mostly primary or secondary school students have been asked to draw a scientist (e.g. Barman, 2009;

Bayri, Koksal, & Ertekin, 2016; Chambers, 1983; Farland, 2006; Finson et al., 1995; Hillman, Bloodsworth, Tilburg, Zeeman, & List, 2014). However, taking into account the potential influence of science teachers on their students' conceptions about science (Christidou, 2011; Türkmen, 2008; see above), it is reasonable to assess (pre-service) teachers' conceptions as well. In line with this, there have been a rising number of studies assessing pre-service teachers' conceptions of scientists by use of the DAST or similar drawing instruments (e.g. McCarthy, 2015; Miele, 2014; Milford & Tippett, 2013; Subramaniam et al., 2013). Some notable differences have been found between groups of pre-service teachers and students. In a study performed by Subramaniam et al. (2013; using the DAST-C), the mythic scientist and indications of technology were not present and indications of danger only rarely drawn. McCarthy (2015), utilising the DAST and DAST-C, found that 'facial hair, mythic stereotypes, secrecy, danger and elderly images were barely represented or absent' (p. 397). Milford and Tippett (2013) deleted the category ethnicity from the DAST-C because they were not able to 'determine the ethnicity of the people in any of the images' (p. 752).

Advantages and disadvantages of drawings and the DAST

What studies in which the DAST or a modified version of it has been used have in common is that they use an external image (drawing) as an indicator for the nature of a person's internal image (conception). In the following section, selected considerations, advantages, and disadvantages will be outlined regarding the use of drawings for the assessment of conceptions in general and within the context of drawing a scientist. Suggestions to solve problematic aspects with the DAST and/or the analysis of the corresponding drawings will be presented.

Ainsworth, Prain, and Tytler (2011) referred to drawings as 'windows into student thinking' (p. 1097). They followed from a literature review that drawings can serve teachers in diagnostic, formative, and summative assessment. Finson and Pederson (2011) emphasised visual images as being a

> well-founded form of data that can be used in a variety of ways that is supported by research across multiple disciplines […]. Clearly visual images are yet another tool researchers can utilize to gain a deeper understanding of the perceptions that students (PK-16) hold of science. (p. 79)

Finson et al. (1995) highlighted that for the DAST no verbal skills are needed, the aspect of social desirability is low, and only a short amount of time is required for assessment.

Van der Veen (2012) points out that drawings create the possibility to 'connect students' internal translations of external experiences through symbolic representations' (p. 365). Hence, considerations can be drawn from symbolism as well. It is shown that 'actions, appearances, and artifacts all carry symbolic significance; greetings, hairstyles, beards, dress styles, and flags may all symbolize socially meaningful matters' (Ball & Smith, 1992, p. 33). Turner (1967; cf. Ball & Smith, 1992) indicates that there are three sorts of meanings which can be assigned to a symbol: the exegetical meaning includes interpretations explicitly given by the informants themselves about the symbol. The operational meaning refers to the application of a symbol in actual practice, that is, what people do and how they interact with the symbol. Finally, the positional or contextual meaning includes the reconstruction of a symbol as part of a whole system. Both meanings need to be observed and interpreted by the researcher (Turner, 1967). It is also highlighted that symbols are essentially polysemic (i.e. they might represent more than one meaning; Ball & Smith, 1992). Visualisation research similarly reveals that external representations (i.e. something that is accessible in the environment) are often connected to additional meanings and associations. For example, a 'flag' refers to the object itself but can also stand for a country, a political mindset or a certain group of persons (Rapp & Kurby, 2008).

In a study of Dove, Everett, and Preece (1999), students aged 9–11 years were prompted to draw a river-basin. The authors discussed if a drawing really represents someone's conception or rather shows the adoption of a common representation. Finson and Pederson (2011) discussed the difference between a conception and the representation of the same conception. They followed that 'a child's drawing will often deviate from the way the child actually perceives the visual world around the child. Consequently, the drawings, by themselves, are of limited value' (p. 76) to reveal their full conceptions of scientists (cf. Golomb, 2002). Ehrlén (2009) interviewed children aged 6–9 about their conceptions of the Earth while they were drawing one. The author implied on the basis of her results that mere drawings from children are not an adequate method to gain reliable information about their conceptions. For example, she found that similar pictures drawn by different students may show different underlying conceptions, which only became apparent by the accompanying interview and not by the drawings themselves. Ehrlén (2009) suggested using drawings together with descriptions about the drawings from the test persons themselves to gain more reliable information.

Concerning the use of the DAST and the DAST-C, Losh et al. (2008) emphasised that in some DAST studies, the participants had to draw several pictures in a row, leading more often to female scientists and scientists belonging to a minority after the first drawing. Likewise, Finson et al. (1995) highlighted that 'students may possess more than one definition of the word scientist' (p. 204; cf. Finson & Pederson, 2011; Maoldomhnaigh & Hunt, 1988). This questions the valid interpretation of single drawings as indicators for people's conceptions of scientists.

Another critical point made by Losh et al. (2008) is associated with 'graphic abilities' (p. 776). For example, although girls drew more female figures, figures made by girls were, in general, more recognisably gendered figures. Especially from first-grade boys, Losh et al. (2008) received many ambiguous gender figures. Thus, they questioned 'the validity of results from earlier studies reporting more clearly gendered figures' made by boys (p. 789). The authors assumed that in earlier studies, the coders were 'forced' to choose between male and female gender and so they recommended a 'probable' or 'indeterminate' gender code.

Miele (2014) reported that despite her prompt to 'draw a scientist doing science' (p. 38), only 24% of her participants drew tools associated with science, hinting at the 'doing science' part. Farland (2006) and Farland-Smith, Finson, Boone, and Yale (2014) recommended a change of the prompt given to test persons. They found that a modification from 'Draw a scientist!' to a more complex prompt, referring explicitly to the three aspects of appearance of a scientist (*appearance*), location a scientist works at (*location*), and scientific activity (*activity*), encouraged students to show more differentiated ideas of what they think a scientist is like. Studies carried out by others (e.g. Finson et al., 1995;

McCarthy, 2015) mostly included the *appearance* and to a minor degree the *location* in their prompt. Farland-Smith et al. (2014) also asked the students to give written information about the sex of the drawn scientist, the *location*, and the *activity*. For the analysis of the drawings, they developed a rubric, which provides a continuum with the drawings labelled as 'cannot be categorised,' 'sensationalised,' 'traditional,' or 'broader than traditional' in regard to *appearance*, *location*, and *activity* (Farland, 2003, 2006; Farland-Smith et al., 2014).

Wentorf, Höffler, and Parchmann (2015) criticised traditional DAST approaches emphasising that the design of these assessments provokes prevailing stigmatisations and prejudices about scientists. However, they do not elaborate this point any further. Wentorf et al. (2015) highlighted that in DAST studies drawing-based inferences often include that students have scientists in mind who are strongly oriented towards manual work (e.g. performing measurements) or who carry out investigative work. The authors developed the *Nature of Scientists* questionnaire including 'scales to analyse students' interests and their self-efficacy next to their views about scientists' activities' (p. 207). Their work was based on the RIASEC model (Holland, 1997) and includes six dimensions, each referring to typical attributes scientists' work can have: Realistic (manual work, e.g. production of medicine), Investigative (analytical, intellectual work, e.g. evaluate results from experiments), Artistic (creative, e.g. develop ideas for new research approaches), Social (caring, e.g. supervise (under)graduates), Enterprising (e.g. raising external funds), and Conventional (administrative, concise, e.g. accounting; Wentorf et al., 2015).

Aim and research questions

Finson et al. (1995) emphasised: 'the scoring of the DAST can be somewhat cumbersome and important stereotypical elements may be omitted during rating and scoring' (p. 201). The present study picks up on this point by aiming to develop a reliable category system for objectively analysing drawings about scientists, which includes the three aspects *appearance*, *location*, and *activity* (Farland, 2006; Farland-Smith et al., 2014). The first research question (RQ) is:

> (RQ 1) To what extent is it possible to objectively categorise pre-service science teachers' drawings about scientists using a deductively developed category system in reference to the *appearance*, the *location*, and the *activity*?

Furthermore, the present article aims to provide information about methodical challenges while assessing and rating drawings of scientists (cf. Farland, 2006; Farland-Smith et al., 2014; Finson, 2002; Losh et al., 2008; Wentorf et al., 2015):

> (RQ 2) What methodical challenges become apparent while assessing and analysing pre-service science teachers' drawings?

Study design

Sample and data collection

Two cohorts of pre-service science teachers (i.e. undergraduate students) participated in the present study. Cohort 1 included 79 pre-service teachers (biology for secondary school: *n* =

61, science for elementary school: $n = 18$; females: $n = 49$, males: $n = 29$, not specified: $n = 1$). After analysing the first cohort's drawings and discussing the results within the group of researchers, the assessment was adjusted (see below) and a second cohort ($N_2 = 101$) of pre-service teachers (biology for secondary school: $n = 101$; females: $n = 61$; males: $n = 38$; not specified: $n = 2$) was asked to draw and to answer a subsequent additional questionnaire. Both surveys took place at the beginning of an introductory course in biology education and prior to any instruction about NOS-related topics. Nevertheless, all of the participants had a second subject (e.g. chemistry, German) and it is not clear to what extent any instruction about NOS aspects was given there. As part of the regular seminar, participants of cohort 1 had about 10 min to draw. Feedback from some of the participants showed that they did not all have enough time to complete the task. Consequently, participants in cohort 2 were given as much time as they needed to fulfil the tasks (approximately 25 min in each course). A few participants were late for the course and had less time because the present researcher had to stop the assessment at some point.

Instruments: mDAST and questionnaire

In order to answer RQ 1, the modified DAST-prompt (mDAST) from Farland (e.g. 2006) was translated into German and adapted for adults: 'Imagine that tomorrow you are going to visit scientists at their place of work. Draw a scientist at work. Label the drawing with a caption that describes the work.'[2] The prompt explicitly addresses the *appearance*, the *location*, and the *activity* in order to receive drawings which can be analysed in all three categories (cf. Miele, 2014). A change of the wording within the prompt can cause a change in the type of drawing (cf. Maoldomhnaigh & Mhaoláin, 1990). Therefore, each participant received a blank paper with the standardised prompt written at the top. After the participants had finished their drawings, they additionally were requested to give written information concerning the sex of the scientist, the *location*, and the *activity* (cf. Farland, 2006). Because of previously experienced interpretation difficulties with some of the *appearance* categories in cohort 1, we added a prompt to describe the scientist himself/herself for cohort 2 (Figure 1).

To answer RQ 2, a questionnaire (Figure 2) was developed on the basis of criticism reported in the literature, on the basis of the discussion of the data of cohort 1, and on

1) Please tick a box:

☐ I drew a woman.
☐ I drew a man.
☐ I drew neither a woman nor a man.

2) Describe **what** the scientist you drew **looks like** (e.g., in reference to age, clothing, hairstyle etc.).

3) Describe **where** the scientist you drew **works** (e.g., inside, outside, in the lab, in the forest etc.).

4) Describe **what** the scientist you drew **is doing**.

Figure 1. mDAST with closed-ended and open-ended items, which were given to the participants while drawing (cf. Farland, 2006; Farland-Smith et al., 2014).

Please answer the following questions.

1. Please evaluate the conditions during the time you were making the drawing.

I had enough time to make the drawing.	☐ yes	☐ no
There were enough materials (e.g., pens).	☐ yes	☐ no
I was successfully able to illustrate my conception in the drawing.	☐ yes	☐ no

Further comments: ____

2. I drew certain characteristics (e.g., lab coat, tubes) so that …

… <u>my</u> conception of a scientist is represented.	☐ yes	☐ no
… my drawing is clearly recognizable as a scientist to <u>others</u>.	☐ yes	☐ no

3. Finish one of the sentences regarding the number of drawn scientists.

I drew <u>one</u> scientist because …

I drew <u>more than one</u> scientist because …

Figure 2. Questionnaire, which asked for the conditions while drawing and reasons certain characteristics were drawn.

the basis of written feedback given by some of the participants from cohort 1. Each participant of cohort 2 was first asked to draw, then to give written information about his/her drawing (Figure 1), and, finally, to fill out the questionnaire (Figure 2).

First, we wanted to find out if limitations of the drawings could be traced back to the assessment time (Figure 2: first item in section 1) because some of the participants in cohort 1 criticised the fact that they had not been given enough time to draw. Next, we assumed that when drawing on white paper, test persons would need to have differently coloured pens to draw scientists of different ethnicities (Losh et al., 2008). We assume that in many studies, invalid indications might have been drawn by inferring that the test persons think a scientist is Caucasian when the colour is not explicitly drawn (Figure 2: second item in section 1). We expected that, similar to children (Losh et al., 2008), adults may have difficulties to express their conceptions in a drawing (Figure 2: third item in section 1). This resulted from the observation that many drawings from cohort 1 included stick figures, which, for example, made it hard to decide whether the scientist is bald or not. Furthermore, we wanted to know whether the drawings represent someone's genuine conception or simply show the adoption of a common representation (Dove et al., 1999; Finson & Pederson, 2011; Figure 2: section 2). Finally, participants of the study should give the reason why they either only drew one or why they drew more than one scientist (Figure 2: section 3). Science is mostly done in small or large groups including different levels and forms of social interactions (Abd-el-Khalick, 2013). It can be seen positively if the test persons draw more than one scientist because it includes a more realistic image of science. At the same time, in many studies (e.g. Laubach, Crofford, & Marek, 2012; McCarthy, 2015; Miele, 2014), Chambers (1983) prompt 'Draw a picture of a scientist' was used and, therefore, it was explicitly asked for only one scientist. We wanted to find out if a prompt that only asks for one scientist would initially provoke the participants to draw only one scientist, irrespective of their actual conception.

Data analysis

For data analysis (RQ 1), we differentiated within the main categories *appearance*, *location*, and *activity* (Farland, 2003, 2006; Farland-Smith, 2012; Farland-Smith et al., 2014). Two coders (first and second authors) coded 10 of the drawings of cohort 1 based on the rubric provided by Farland-Smith. The coders agreed that the rubric had been too rough to pick up conceptions of scientists and their working field in a differentiated manner. The rubric, including a continuum represented by four categories ('cannot be categorised,' 'sensationalised,' 'traditional,' and 'broader than traditional'), gives the impression that students' conceptions are rated with regard to a level of adequacy. The issue of adequacy, however, is difficult to grasp because 'science is so rich and so dynamic and scientific disciplines are so varied that there seems to be no set of features that is common to all of them and shared only by them' (Irzik & Nola, 2011, p. 591). It seems that it is not feasible to judge someone's drawing of a scientist to be appropriate or not. Rather, learners' conceptions need to be descriptively reported in order to serve for the construction of learning environments (cf. Duit et al., 2012). We assigned sub-categories to the three main categories, which were adapted from Finson et al. (1995), McCarthy (2015), Milford and Tippett (2013), and Wentorf et al. (2015). On this basis, we deductively developed a category system (cf. Mayring, 2010). Refinement, evaluation, and application of the category system (Tables 1–3) were achieved through the following steps:

(1) Refinement of the category system: Experts in biology education research ($N = 9$) commented on the category system independently in a written form. The category system was subsequently discussed with all experts and modified accordingly.
(2) Evaluation of the category system: A heterogeneous sub-sample of the drawings and the corresponding written answers ($n = 21$ from cohort 1) were independently coded by the first and second authors. The codings were compared and discussed. The category system and the coding guidelines were modified and refined. For example, the category *unrealistic* was added to *activity* to take account of dangerous or mythic activities, which were not included in the study of Wentorf et al. (2015). Also, we changed some of the sub-category names within *activity*, which we adopted from Wentorf et al. (2015), because we found that they did not precisely describe the content of the category (*artistic* to *artistic – engineering*; *networking* to *social – scientific exchange*; Table 3).
(3) Application of the category system for cohort 1: All drawings as well as all written answers of cohort 1 ($N_1 = 79$) were coded by the first author and, additionally, by one of the third to fifth authors. If the drawing and the written answer were contradictory, the latter was given priority. Cohen's kappa (κ) was calculated and differing

Table 1. Absolute frequency and interrater reliability κ for the sub-categories of *appearance*.

		Cohort 1 ($N = 79$)				Cohort 2 ($N = 101$)			
Category	Category description	Absolute frequency			κ	Absolute frequency			κ
Sex	Closed item	Female	Male	Neuter	–	Female	Male	Neuter	–
		17	42	20		23	39	39	
Number of scientists	One scientist or more than one scientist (drawn students are not counted)	One	Two or more		0.79	One	Two or more		1.0
		76	3			86	15		
Age	Young: up to 30 years Middle-aged: between 30 and 50 years Old: 50 years and older Ageless: age is not seen as important or covers all age groups	–			–	Young	Middle	Old	0.92
						20	40	6	
						Ageless	Unratable		
						15	20		
Clothes	Lab coat Everyday clothes Both: wearing lab coat and everyday clothes Not important: clothes are not seen as important	Lab coat			0.70	Lab coat	Everyday clothes	Both	0.83
		33				56	9	12	
						Not important	Unratable		
						11	13		
Protective clothing	Clothing that serves as protection, e.g. from dangerous substances (except lab coat, glasses)	2			0.79	12			0.82
Glasses	Scientist wears glasses	38			0.98	52			1.0
Headgear	Hat, cap, etc.	3			1.0	5			0.79
Beard	Scientist has a beard	7			0.84	5			1.0
(Half-)bald head	Bald or half-bald head (even if it is a stick figure) Not important: hair style is not seen as important	(Half-)bald head			0.57	(Half-)bald head	Not important		0.68
		19				9	10		
Fictive	Scientist(s) from movies or literature	1			1.0	1			*

*Cannot be computed; the frequencies of the category *age* and some characteristics of the categories *clothes* and (*half-*)*bald head* could not be given in cohort 1 due to coding difficulties.

Table 2. Absolute frequency and interrater reliability κ for the sub-categories of *location*.

Category	Category description	Cohort 1 (N = 79) Absolute frequency		κ	Cohort 2 (N = 101) Absolute frequency		κ
Surroundings	Inside: e.g. lab, office Outside: e.g. sea, forest Both: locations inside and outside were drawn	Inside 71 Both 3	Outside 4 Unratable 1	0.95	Inside 82 Both 13	Outside 6	0.97
Technology	E.g. computer, rockets	14		0.95	37		0.89
Knowledge	E.g. books, formulas, scientific terms, handwriting on a blackboard in the context of knowledge transfer	29		0.75	49		0.90
Research instruments	Pipettes, magnifying glass, Bunsen burner, test tubes, etc.	68		0.86	92		0.83
Danger	E.g. lots of smoke, explosions, monsters	4		0.74	4		0.74

Table 3. Absolute frequency and interrater reliability κ for the sub-categories of *activity*.

Category	Category description	Cohort 1 (N = 79) Absolute frequency	κ	Cohort 2 (N = 101) Absolute frequency	κ
Inquiry	Manual and/or analytical, task-oriented activities, e.g. taking measurements, using the microscope, observing, experimenting, reading literature	71	0.49	97	0.74
Artistic – engineering	Creative activities, e.g. developing inventions, measuring instruments	3	0.50	0	*
Social – teaching	Knowledge transfer, e.g. teaching/supervising (PhD) students	3	1.0	1	1.0
Enterprising	E.g. raising funds, managing research groups, writing articles	5	0.88	5	0.32
Conventional	Administrative activities, e.g. accounting	–	–	1	*
Social – scientific exchange	Networking activities, e.g. conferences, meeting with colleagues of other departments	2	1	14	0.92
Unrealistic	Fictive activities, e.g. creation of secret substances	3	0.79	1	1.0

*Cannot be computed.

codings were discussed until consensus was reached (consensual coding; cf. Schmidt, 2010). Selective modifications of the coding guidelines were made. Thus, we combined the categories *realistic* and *investigative* under the name of the latter because often no consensus could be reached about a differentiation in the drawings. For example, in Figure 3 the *activity* was coded as being investigative because the written information about the *activity* was 'analysing results.' However, it was not clear if the drawing itself also included a realistic perspective (i.e. manual work) because the laboratory equipment drawn (test tubes, flask) and the goggles gave the impression that the scientist was also performing some manual laboratory work.

(4) Application of the category system for cohort 2: Because we were not able to categorise the *appearance* of the drawn scientist regarding some characteristics (e.g. age) in cohort 1, we added the prompt to describe the scientist himself/herself (see above), which led to the addition of the category *age*, the extension of the category *lab coat* to *clothes*, and the refinement of the category (*half-*)*bald head* (Table 1). The drawings and the corresponding written answers of cohort 2 were coded as described for cohort 1 (step 3).

Figure 3. Drawing by pre-service teacher [p45] from cohort 1.
Note: [pX]= person number as assigned after the assessment.

With the questionnaire, we aimed at receiving evidence for methodical challenges (RQ 2) which had been perceived by the participants. The absolute frequencies of chosen answer options were counted for each of the closed-ended items (Figure 2: sections 1 and 2), the written answers were inductively categorised and a name as well as a description of each category was given.

Results

Results for RQ 1: category system

Except for the sub-category *enterprising* for cohort 2, we can report an acceptable to a very good interrater agreement for each sub-category (Tables 1–3: $0.49 < \kappa < 1.0$ for cohort 1; $0.68 < \kappa < 1.0$ for cohort 2; cf. Wirtz & Caspar, 2002). Concerning the *appearance* of a scientist, the most frequent characteristics are solitary-working, middle-aged, lab coat, and glasses (e.g. Figures 4 and 5; Table 1). Most drawings set the *location* to be inside a room with indications of knowledge and research instruments (e.g. Figures 3–7; Table 2). As for *activities*, almost all of the participants drew or named inquiry activities (e.g. Figures 3–7) and included other activities only to a minor degree (e.g. Figures 5 and 7; Table 3).

Some of the participants in cohort 2 wrote that the age, the appearance, or some features of the appearance (e.g. hair style) are not important. We added 'not important' as one possible characteristic of the categories *clothes* and (*half-*)*bald head* and 'ageless' as a characteristic of the category *age*. Between 10 and 15 of the drawings from cohort 2 were rated accordingly (Table 1).

Figure 4. Drawing by pre-service teacher [p26] from cohort 2.

Figure 5. Drawing by pre-service teacher [p50] from cohort 2.

Figure 6. Drawing by pre-service teacher [p74] from cohort 1.

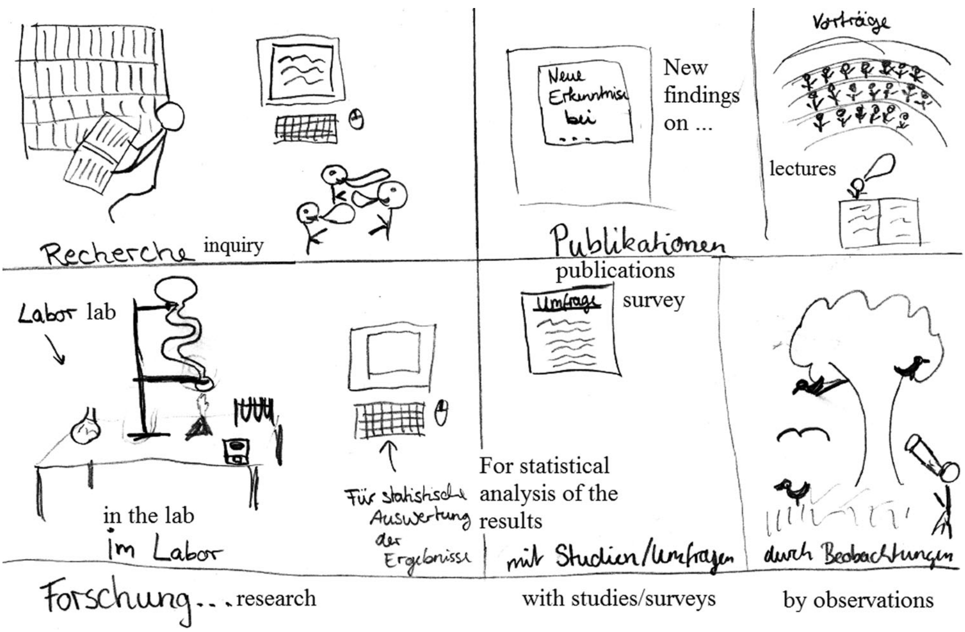

Figure 7. Drawing by pre-service teacher [p23] from cohort 2.

For the category *fictive figure*, which extends the mythic stereotype (McCarthy, 2015; Milford & Tippett, 2013), we had one drawing in cohort 1 showing the comic-strip hero Iron Man [p17] and one drawing in cohort 2 showing a scientist with a robot arm [p66].

Table 4. Absolute frequency to the items referring to the conditions while drawing.

	Yes	No	Unratable
I had enough time to make the drawing.	82	15	4
There was enough material (e.g. pens) available.	73	26	2
I have succeeded to show my conception graphically.	67	31	3

Table 5. Absolute frequency to the items referring to the aim of the drawings.

		For others			
		Yes	No	Unratable	Σ
For me	Yes	53	11	5	69
	No	17	6	0	23
	Unratable	5	1	3	9
	Σ	75	18	8	101

Note: For me = participant drew certain characteristics to show his/her own conception; for others = participants drew certain characteristics to make drawing clearly recognisable as a scientist to others.

Results for RQ 2: methodical challenges

The participants in cohort 2 were asked to answer an additional questionnaire (Figure 2; Table 4). Almost one-third of the participants ($n = 31$) gave further comments, from which 14 referred to their lack of drawing skills and 7 referred to the lack of time and/or material available (e.g. 'I only had a fine-tipped pen and little time' [p38]). There were two comments about the lack of detail in the drawings (e.g. 'I did not draw too precisely' [p18]). Finally, there were nine rather diverse comments such as 'Science is a too large subject area to summarize it in one drawing (referring to their work)' [p8], 'maybe I should have drawn several scenes, e.g. next to a field experiment, a lab experiment … ' [p25], or 'strongly influenced by experience' [p50].

Most of the participants both drew their own conception and wanted to make their scientist recognisable for others (Figure 2; Table 5).

We received 10 different reasons why one scientist was drawn and 7 different reasons for drawing several scientists (Table 6).

Discussion

Category system: appearance, location, *and* activity *(RQ 1)*

In traditional DAST studies, the participants' conceptions of the *appearance* and to a minor degree the *location* have been assessed and analysed. In the present approach, the *appearance*, *location*, and *activity* were assessed in conjunction (cf. Farland-Smith, 2012), which affected the assessment as well as the analysis of the pre-service teachers' conceptions. The deductively developed category system (e.g. McCarthy, 2015) was inductively extended by means of the drawings (see *Data analysis*). With regard to Farland (2006) and Farland-Smith et al. (2014), we added a closed-ended item in order to assess the gender of the drawn scientist. Participants had to decide between male, female, and neuter, which is different to reports from other studies. In these studies, either the presence of a male scientist was assessed (e.g. McCarthy, 2015) or it was differentiated between male, female, and indeterminate (e.g. Miele, 2014). It is noteworthy that

Table 6. Reasons given by the pre-service teachers why one or more than one scientist was drawn.

Category	Example	*n*
Reasons for one scientist: I drew one scientist because …		
Depicts his/her own conception	… for me it is a profession that is based on less collaboration than others (not a social profession). [p29]	38
Task demanded to draw only one scientist	… the task was that you should only draw one, otherwise I would have drawn several. [p1]	22
Lack of time	… I only had time to draw one. [p65]	9
Lack of drawing ability	… I cannot draw people. One drawing disaster is enough. [p69]	9
Lack of motivation	… I had no desire to draw several. [p26]	7
Lack of space	… there would not have been enough space to draw several […]. [p30]	2
Intention to only show one conception	… I only wanted to illustrate one idea. [p16]	3
Self-portrait	… this is about showing my own reference point, meaning my work. [p58]	3
One scientist at different locations/ doing different things	… I wanted to make clear that one and the same scientist works at several places. [p53]	2
Not clearly interpretable	… I wanted to ('male profession'). [p8]	3
Reasons for more than one scientist: I drew more than one scientist because …		
Scientists work together	… I think that you can do better research in a team, scientists exchange information with each other and often groups of several people do research on one question/hypothesis. [p23]	6
Science, its actors and activities are diverse	… science is diverse and there is no 'one' scientist. [p21]	3
There are several scientists working in a lab	… there are always several scientists in a lab. [p31]	2
Not only one sex should be shown	… I wanted to make clear that there are female as well as male scientists. [p45]	2
Decisions are not made only by one person	… negotiations are conducted at a climate summit. Decisions are not made by only one person. [p80]	1
The task demanded to draw several scientists	… in the task it said 'scientists' and I associated this with plurality. [p90]	1
To make the drawing clearer	… it makes the drawing clearer. [p79]	1

Note: *n* = number of participants, who gave this reason; [pX] = person number as assigned after the assessment.

about one-third of all participants (Table 1) chose *neuter* and, thus, we suggest including this category permanently in DASTs.

In cohort 2, we also asked for the overall appearance of the drawn scientist because it was difficult to rate the drawings in cohort 1 in some categories. For the category *(half-) bald head*, the interrater reliability was $\kappa = 0.57$ in cohort 1, which is rather low compared to most of the other categories (Table 1). This was traced back to the issue that often stick figures with no hair had been drawn and it was not clear whether this was intended to show a bald headed person or not. Some of the drawn figures included hair only to some degree. It was not clear if a half-bald head was imagined or whether it resulted from a lack of drawing skills. In cohort 2, Kappa was higher for this category ($\kappa = 0.68$) than in cohort 1 which can be justified with the additional question (Figure 1).

It is notable that because of the additional prompt asking for a description of the scientists' appearance (Figure 1), some drawings were rated as 'not important' in the categories *(half-)bald head* and *clothes* as well as rated as 'ageless' in the category *age* (Table 1). Such ratings cannot be found in other DAST studies (e.g. McCarthy, 2015; Subramaniam et al., 2013) but they reveal an important aspect. At least 10% of the participants from cohort 2 highlighted that at a minimum some aspects of a scientist's appearance are not important. Without explicitly asking about the *appearance*, this information would have been lost. For some pre-service teachers something else is more important than how the scientist looks, for example, the kind of work he or she is doing, as the following statement confirms: 'I drew stick figures for which neither age nor sex, clothes, hair style … are identifiable. I focused on the activity a scientist does' [p23].

Within the category *appearance*, we did not find any assignment to the sub-category *popular figure*. Contrarily, Miele (2014) reported that in her study many undergraduates' 'images represented real persons such as Albert Einstein, television science personalities, or former high school and college science teachers that fit the common stereotypes' (p. 37). In Miele's (2014) study, the participants verbally self-evaluated their drawings. Thus, less interpretation difficulties exist, but more time with the participants and their willingness to not only draw but also to analyse their drawings is needed. Additionally, this procedure would lead to a renunciation of anonymity.

Studies in which the DAST-C was used include several characteristics that indicate a mythic or even dangerous image of scientists. Some of these features clearly belong to the *appearance* of a scientist (wild hair) and others could be, in our terms, part of the *activity* and the *location* (indications of danger, indications of secrecy, the mythic stereotype; McCarthy, 2015; Miele, 2014). Results referring to the categories *fictive figure* ($n = 2$; Table 1), *indications of danger* ($n = 4$; Table 2), and *unrealistic activities* ($n = 3$; Table 3) are in line with the results of McCarthy (2015) and Subramaniam et al. (2013), who reported that the mythic stereotype and indications of secrecy or danger were not or only barely represented in the drawings of the participating pre-service teachers. Based on our results, we agree 'that the items [i.e. categories] of secrecy and mythic stereotypes could be eliminated […] for older students' (McCarthy, 2015, p. 408).

Concerning *activity*, a comparison with the study of Wentorf et al. (2015) reveals some commonalities as well as differences to our results. Wentorf et al. (2015) reported that the secondary school students who took part in their study placed a high significance on realistic, investigative (here both combined to *investigative*), artistic (here *artistic – engineering*), and networking (here *social – scientific exchange*) activities and they viewed enterprising activities as less significant. We, on the other hand, received a high number of drawings and/or corresponding explanations merely with regard to investigative activities. The remaining activities were present only to a small degree (Table 3). A reason for this could be that the pre-service teachers drew only what they think is the most prominent activity scientists perform. In a second drawing, an alternative activity might be shown. Likewise, Wentorf et al. (2015) criticised that by using drawings and corresponding interviews, it is primarily conceptions of realistic or investigative activities that are being provoked. Following this, a valid assessment of all conceptions an individual might hold by means of only one drawing should be seen as critical (cf. Finson et al., 1995; Finson & Pederson, 2011; Losh et al., 2008). This is supported by some of the drawings in which the pre-service teachers divided the sheet of paper in several sections and drew more than one picture without any prompt (e.g. Figure 7).

Additionally, we inductively developed the category *unrealistic* as another sub-category of *activity*, which can be explained by the use of an open-ended instead of the closed-ended instrument by Wentorf et al. (2015). This can be seen as a general strength of qualitative analyses: 'data are based on the participants' own categories of meaning' (Johnson & Onwuegbuzie, 2004, p. 20).

Methodical challenges with the DAST

First, some remarks have to be made in regard to the above-outlined considerations of symbolism (cf. Ball & Smith, 1992; Turner, 1967). It can be assumed that the drawings include symbols with corresponding meaning(s), which need a careful interpretation. For example, Table 1 shows that in sum 51 drawings included some sort of a technological artefact. In most cases, computers or laptops were shown (e.g. Figures 5–7). To achieve information about the exegetical meaning (see above; Turner, 1967), participants' answers need to be considered. For example, two participants of cohort 2 who included a computer or laptop in each their drawings wrote: 'analysis – collate results' [p60], 'the man writes e-mails and publications' [p89]. Consequently, different activities are connected with the symbol of a computer (categories: inquiry, social-scientific exchange). Without the additional prompt, this information would not be available and it is suggested to always gather additional descriptions about the drawings from the test persons within DAST studies (see also Ehrlén, 2009). It must be noted that not every participant explained why there was a computer in the drawing. To interpret the exegetical meanings of every drawn 'symbol' (lab coat, beard, glasses, etc.), the participants would need to explain each single symbol in their drawings. The operational and contextual meanings of the drawn computers cannot be reconstructed due to missing information. By means of example, the operational meaning could give information about which ways a participant uses a computer and, therefore, which attributes or activities he or she associates with it. The contextual meaning would include how the symbol of a computer is understood within the cultural context. Contextual information could be especially interesting for DAST studies in which the cultural influences on students' or pre-service teachers' drawings are analysed (e.g. Walls, 2012).

Within symbolism as well as visualisation research (see above), the polysemous meaning of a symbol is highlighted (Ball & Smith, 1992; Rapp & Kurby, 2008), and this becomes important in the present study as well. Exemplarily, the drawing of [p50] will be discussed (Figure 5). Next to the scientist a cell phone is ringing. This could have different meanings: one could be that the scientist is receiving a call; another could be that an alarm clock is ringing. The meaning could be interpreted in two ways. Since the scientist seems to be focused on his work (interpreted from the concentrated facial expression of the scientist), he could have forgotten a meeting date or appointment and receives a call from someone, who wants to remind him. The pre-service teacher could indicate that a scientist is forgetful, maybe unorganised. Also, this would fit to the written description of the appearance of the scientist: '[…], hair style: a little messy, […].' Another interpretation might be that the scientist is an important person and receives many phone calls on a regular basis. The written answer of [p50] also includes: 'The scientist works in his office, where he spends most of his time. Lab and field work is mostly done by his assistants, interns and students.' The ringing phone could mean that he needs to communicate a lot with other researchers in his research group. Communicative and interpersonal skills could be associated with a scientist by the pre-service teacher. The problem of polysemy is illustrated by this example. Both forgetfulness and communicative skills of scientists could be symbolised by the ringing cell phone (there are probably even more meanings), and these are very different images of scientists. Other symbols within the drawing of [p50] that are polysemous are the bird on the bookshelf (symbol for biology, use of model organisms in research, or just decoration), the glasses (much time spending on the computer and/or reading, high mental abilities, scientist is a 'nerd') or the large amount of books and notes (high literacy, heavy workload). The above examples show that a data analysis of DAST-drawings only by content analysis is not sufficient to validly assess pre-service teachers' (and students') conceptions about scientists and the working field (cf. Ball & Smith, 1992).

The following methodical problems are based on the participants' answers to the questionnaire or became apparent while analysing the drawings. Almost one-third of the pre-service teachers answered in the questionnaire that they were not able to depict their idea in the drawing (Table 4). This shows that graphic abilities influence the outcome of the drawing task (Losh et al., 2008). It can be assumed that good drawing abilities could lead to a better, although not necessarily clearly interpretable (see discussion above for Figure 5), representation of conceptions by means of drawings. This is also supported by 14 additional comments from pre-service teachers addressing a lack of drawing skills (see above: Results for RQ 2). We expect that this becomes even more important when investigating conceptions of younger students with probably even less developed drawings skills.

In alignment with other drawing studies (Dove et al., 1999; Finson & Pederson, 2011; Golomb, 2002), we presumed that the participants will draw certain stereotypical characteristics of scientists (e.g. lab coat, glasses) to show their own conception, but also to make their own drawing clearly recognisable as a scientist to other people. Results show that more than half of the pre-service teachers had both in mind and 17 even drew certain characteristics to make their drawing clearly recognisable as a scientist to others and not to show his/her own conception (Table 5). The aim of DAST studies is to elicit the participants' own conceptions. To obtain corresponding data, we follow that there must be bigger emphasis in the prompt to advise participants to depict their own conception rather than showing common representations (cf. Dove et al., 1999; Finson & Pederson, 2011). However, such a prompt might lead to drawings in which the participants paid explicit attention *not* to draw a stereotypical image; although, such an image would correspond to their conception. Consequently, formulating an adequate prompt could be challenging and must be performed carefully. Alternatively, non-reactive methods could be used to collect ideas, which would lead to less distorted data (Fritsche & Linneweber, 2006; for unobtrusive measures cf. Webb, Campbell, Schwartz, & Sechrest, 1966). One example is the elicitation of conceptions by means of (focus) group discussions, in which a natural conversation/discussion situation is generated (cf. Ryan, Gandha, Culbertson, & Carlson, 2014).

In line with Farland (e.g. 2006), we added three questions in cohort 1 asking the participants explicitly for the sex, the *location*, and the *activity* (see section *Instruments: mDAST and Questionnaire*). We found this procedure helpful for analysing the drawings. For example, in many drawings, the sex of the scientists could not have been determined otherwise (cf. Losh et al., 2008). In cohort 2, we added another question asking for a description of the drawn scientists' overall appearance, because for the drawings of cohort 1, it was sometimes not clear if the scientist was wearing a lab coat or another type of clothing (e.g. a dress). Another example concerns the age of the drawn scientists, which is seen as an indicator for stereotypical thinking: 'The image of a middle-aged man […] is a widely held stereotype of a scientist' (Milford & Tippett, 2013, p. 759). We had to delete this category for cohort 1 because we were not able to determine the age in most of the drawings. An indicator for the older age of a male scientist could be the existence of a bald or half-bald head. As outlined before, the rating for this category proved to be difficult in many drawings of cohort 1 as well. We implicate from these methodical problems that an open-ended item concerning the appearance of a scientist needs to be included, as we have done for cohort 2, or questions with a closed-ended instrument referring specifically to some characteristics of the scientists, such as the age, need to be integrated.

Another problematic issue concerns the number of drawn scientists. Only 18 out of 180 pre-service teachers drew more than one scientist (Table 1). This is in line with other studies, which showed that the test persons often see scientists as single workers (e.g. Cakmakci et al., 2011). Considering the results for the questionnaire, this is partly because the pre-service teachers thought that scientists work on their own, but also because the prompt demanded to draw only one scientist (Table 6). Miele (2014) discussed to what extent the prompt 'Draw a scientist' is taken literally by the test persons and does not represent their actual conception. We suggest modifying the prompt and leaving it to the test persons to draw one or more than one scientist, for example: 'Imagine that tomorrow you are going to visit scientists at their place of work. Draw one or more than one scientist at work. Label the drawing with a caption that describes the work.' Another possible reason for the high amount of drawings, which included only one scientist, is that there was not enough time to draw (Tables 4 and 6). This can be seen as a particular limitation of the present study. Some pre-service teachers also did not feel comfortable to draw more than one person because of their self-perceived lack of drawing skills or they indicated that they were not motivated to draw more than one person (Table 6). In sum, most of the reasons (Table 6) do not match the conception of scientists working on their own. However, traditionally, the drawings might be interpreted as such.

Referring to *appearance*, ethnicity was often rated in previous studies (e.g. DAST-C; Finson et al., 1995). McCarthy (2015) stated in regard to her study: 'A conspicuous finding was the absence of diversity in the drawings. Even though there were seven Black participants in the study, only one Black scientist appeared' (p. 407). We deleted the category ethnicity for two reasons: First, none but one pre-service teacher (Figure 6) used any colours for his/her scientist(s). In the exceptional case, the face was coloured yellow, but did not include any features and no other part of the body was visible. We assume that this was done because otherwise the face would not have been recognisable as a face at all rather than to draw yellow skin colour intentionally. Second, there was a consensus among the experts that someone who is not colouring his/her scientist does not necessarily want to draw a Caucasian. This has been interpreted differently in previous studies (e.g. Hillman et al., 2014; McCarthy, 2015). We acted in agreement with Milford and Tippett (2013), who were not able to determine the ethnicity in any drawing and thus deleted the category as well. Results for the questionnaire show that about three quarters of the participants (Table 4) indicate that there was no lack of material to fulfil the task. In order to draw a scientist with a different ethnicity you may have to colour the figure. This seems unlikely because in most of the drawings only one pen was used. We assume that the participants do not care as much about the ethnicity of the scientist, which is supported by at least some statements in which the participants noted that the appearance is not important (see above). This could be investigated further by interviews/thinkaloud protocols or by giving different coloured pens and papers to draw on and then analyse if participants use a white or skin-coloured pen to symbolise a Caucasian.

Conclusion

In 1983, Chambers highlighted that 'because DAST does not rely on verbal response, it can be utilized at an earlier age than other attitude measuring tests' (p. 264). Considering the present results and above-outlined discussion, the use of the DAST or similar instruments for the assessment of older students' or even adults' conceptions should be seen critically. It has been discussed (Finson et al., 1995; Finson & Pederson, 2011) that the DAST and the DAST-C may not reveal the complete array of conceptions students have of scientists and that interviews or narratives as supporting measures to the drawings may rather be used (e.g. Barman, 2009). Likewise, in regard to considerations from symbolism, the above presented discussion shows that the drawings reveal much more than what is portrayed on a piece of paper. As a consequence, within DAST and most likely within drawing studies in general, additional explanations need to be given by the participants.

As outlined above, the prompt given to DAST participants is another critical issue within DAST studies. Results should be evaluated while keeping in mind a possible influence of the given prompt. Alternatively, non-reactive methods such as (focus) group discussions could be a form of measurement. Including additional verbal measures, such as (focus) group discussions or interviews, would give more information, but is very time-consuming in return (Johnson & Onwuegbuzie, 2004). Therefore, only small populations could be assessed. We expect that the development of an open-ended or closed-ended instrument, which considers not only the *activity* (Wentorf et al., 2015), but also the *appearance* and *location*, would be useful to assess older students' and adults' conceptions of scientists and their work. In this study, several methodical problems that limit the interpretations of the drawings became obvious. The argument that writing skills are not needed in drawing-assessments is only important when assessing primary school students or even younger children and is not meaningful when assessing older students or adults.

Two directions for further research can be proposed based on the findings of the present study. First, the symbolic meaning of certain objects within DAST-drawings should be analysed to achieve more in-depth knowledge about present conceptions about scientists and their work. Second, it can be concluded that an open-ended or closed-ended instrument might be more appropriate to validly assess test persons' conceptions of scientists and their work. The category system presented in this study as well as results from a symbolic analysis could form a basis for the analysis of answers given in an open-ended instrument. Thus, we suggest developing and evaluating such an instrument.

Notes

1. For a detailed discussion about the integration of NOS in Shulmans' classification, see van Dijk (2014).
2. It ought to be noted that there is a gender-specific difference for the word 'scientist' in German. Thus, the following prompt was originally given: 'Stellen Sie sich vor, Sie besuchen

morgen Naturwissenschaftler/innen an ihrem Arbeitsplatz. Zeichnen Sie eine/n Naturwissenschaftler/in bei der Arbeit. Beschriften Sie die Zeichnung mit einem Titel, der diese Arbeit beschreibt.'

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Bianca Reinisch http://orcid.org/0000-0001-9788-3388

References

Abd-el-Khalick, F. (2013). Teaching with and about nature of science, and science teacher knowledge domains. *Science & Education*, *22*, 2087–2107.

Ainsworth, S., Prain, V., & Tytler, R. (2011). Science education. Drawing to learn in science. *Science*, *333*, 1096–1097.

Ball, M. S., & Smith, G. W. H. (1992). *Analyzing visual data*. Newbury Park, CA: Sage.

Barman, C. R. (2009). Using visual data to obtain students' perceptions of scientists and studying science. In J. E. Pedersen & K. D. Finson (Eds.), *Visual data. Understanding and applying visual data to research in education* (pp. 125–131). Rotterdam: Sense.

Bayri, N., Koksal, M. S., & Ertekin, P. (2016). Investigating gifted middle school students' images about scientists. *Science Education International*, *27*, 136–150.

Burton, E. P. (2013). Student work products as a teaching tool for nature of science pedagogical knowledge: A professional development project with in-service secondary science teachers. *Teaching and Teacher Education*, *29*, 156–166.

Cakmakci, G., Tosun, O., Turgut, S., Orenler, S., Sengul, K., & Top, G. (2011). Promoting an inclusive image of scientists among students: Towards research evidence-based practice. *International Journal of Science and Mathematics Education*, *9*, 627–655.

Campbell, T., Schwarz, C., & Windschitl, M. (2016). What we call misconceptions may be necessary stepping-stones toward making sense of the world. *Science and Children*, *53*(7), 28–33.

Capps, D., & Crawford, B. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education*, *24*, 497–526.

Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science Education*, *67*, 255–265.

Christidou, V. (2011). Interest, attitudes and images related to science: Combining students' voices with the voices of school science, teachers, and popular science. *International Journal of Environmental & Science Education*, *6*, 141–159.

Dove, J. E., Everett, L. A., & Preece, P. F. W. (1999). Exploring a hydrological concept through children's drawings. *International Journal of Science Education*, *21*, 485–497.

Duit, R., Gropengießer, H., Kattmann, U., Komorek, M., & Parchmann, I. (2012). The model of educational reconstruction. In D. Jorde & J. Dillon (Eds.), *Science education research and practice in Europe* (pp. 13–37). Rotterdam: Sense.

Ehrlén, K. (2009). Drawings as representations of children's conceptions. *International Journal of Science Education*, *31*, 41–57.

Farland-Smith, D. (2012). Development and field test of the modified draw-a-scientist test and the draw-a-scientist rubric. *School Science and Mathematics*, *112*, 109–116.

Farland-Smith, D., Finson, K., Boone, W. J., & Yale, M. (2014). An investigation of media influences on elementary students representations of scientists. *Journal of Science Teacher Education*, *25*, 355–366.

Farland, D. (2003). *The effect of historical, non-fiction, trade books on third grade students' perceptions of scientists* (Dissertation). Lowell, MA: University of Massachusetts.
Farland, D. (2006). The effect of historical, nonfiction trade books on elementary students' perceptions of scientists. *Journal of Elementary Science Education*, *18*(2), 31–47.
Finson, K. D. (2002). Drawing a scientist: What we do and do not know after fifty years of drawings. *School Science and Mathematics*, *102*, 335–345.
Finson, K. D., Beaver, J. B., & Cramond, B. L. (1995). Development and field test of a checklist for the draw-a-scientist test. *School Science and Mathematics*, *95*, 195–205.
Finson, K., & Pederson, J. (2011). What are visual data and what utility do they have in science education? *Journal of Visual Literacy*, *30*, 66–85.
Fritsche, I., & Linneweber, V. (2006). Nicht-reaktive Verfahren[Non-reactive methods]. In F. Petermann & M. Eid (Eds.), *Handbuch der psychologischen Diagnostik* (pp. 203–210). Göttingen: Hogrefe.
Golomb, C. (2002). *Child art in context: A cultural and comparative perspective*. Washington, DC: American Psychological Association.
Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York, NY: Teachers College.
Günther, S. L., Fleige, J., Upmeier zu Belzen, A., & Krüger, D. (2017). *Using the case method to foster preservice biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK) related to models and modeling*. Manuscript submitted for publication.
Hillman, S. J., Bloodsworth, K. H., Tilburg, C. E., Zeeman, S. I., & List, H. E. (2014). K-12 students' perceptions of scientists: Finding a valid measurement and exploring whether exposure to scientists makes an impact. *International Journal of Science Education*, *36*, 2580–2595.
Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources.
Irzik, G., & Nola, R. (2011). A family resemblance approach to the nature of science for science education. *Science & Education*, *20*, 591–607.
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research. A research paradigm whose time has come. *Educational Researcher*, *33*, 14–26.
KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der BRD]. (Ed.). (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*[Biology education standards for the Mittlere Schulabschluss]. Munich: Wolters Kluwer Deutschland. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Biologie.pdf
Krell, M., Koska, J., Penning, F., & Krüger, D. (2015). Fostering pre-service teachers' views about nature of science: Evaluation of a new STEM curriculum. *Research in Science & Technological Education*, *33*, 344–365.
Laubach, T. A., Crofford, G. D., & Marek, E. A. (2012). Exploring native American students' perceptions of scientists. *International Journal of Science Education*, *34*, 1769–1794.
Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–880). Mahwah, NJ: Lawrence Erlbaum.
Lederman, N. G., & Lederman, J. S. (2014). Research on teaching and learning of nature of science. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 600–620). New York, NY: Routledge.
Losh, S. C., Wilke, R., & Pop, M. (2008). Some methodological issues with "draw a scientist tests" among young children. *International Journal of Science Education*, *30*, 773–792.
Maoldomhnaigh, M. Ó., & Hunt, Á. (1988). Some factors affecting the image of the scientist drawn by older primary school pupils. *Research in Science and Technological Education*, *6*, 159–166.
Maoldomhnaigh, M. Ó., & Mhaoláin, V. N. (1990). The perceived expectation of the administrator as a factor affecting the sex of scientists drawn by early adolescent girls. *Research in Science & Technological Education*, *8*, 69–74.
Mayring, P. (2010). *Qualitative Inhaltsanalyse*[Qualitative content analysis]. Weinheim: Beltz.

McCarthy, D. (2015). Teacher candidates' perceptions of scientists: Images and attributes. *Educational Review*, *67*, 389–413.
Mead, M., & Métraux, R. (1957). Image of the scientist among high-school students. A pilot study. *Science*, *126*(3270), 384–390.
Miele, E. (2014). Using the draw-a-scientist test for inquiry and evaluation. *Journal of College Science Teaching*, *43*(4), 36–40.
Milford, T. M., & Tippett, C. D. (2013). Preservice teachers' images of scientists: Do prior science experiences make a difference? *Journal of Science Teacher Education*, *24*, 745–762.
NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies.
Pansegrau, P. (2008). Stereotypes and images of scientists in fiction films. In B.-R. Hüppauf & P. Weingart (Eds.), *Science images and popular images of the sciences* (pp. 257–266). New York, NY: Routledge.
Rapp, D. N., & Kurby, C. A. (2008). The 'ins' and 'outs' of learning: Internal representations and external visualizations. In J. Gilbert, M. Reiner, & M. B. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 29–52). New York, NY: Springer.
Reinisch, B., & Krüger, D. (2016). Pre-service biology teachers' conceptions about the tentative nature of theories and models in biology. *Research in Science Education*. doi:10.1007/s11165-016-9559-1
Ryan, K. E., Gandha, T., Culbertson, M. J., & Carlson, C. (2014). Focus group evidence: Implications for design and analysis. *American Journal of Evaluation*, *35*, 328–345.
Schmidt, C. (2010). Auswertungstechniken fürLeitfadeninterviews [Analysing technique for guided interviews]. In H. Boller, B. Friebertshäuser, A. Langer, A. Prengel, & S. Richter (Eds.), *Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft* (pp. 473–486). Weinheim: Juventa.
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*, 4–14.
Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*, 1–23.
Subramaniam, K., Esprívalo Harrell, P., & Wojnowski, D. (2013). Analyzing prospective teachers' images of scientists using positive, negative and stereotypical images of scientists. *Research in Science & Technological Education*, *31*, 66–89.
Treagust, D. F., & Duit, R. (2008). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, *3*, 297–328.
Türkmen, H. (2008). Turkish primary students' perceptions about scientist and what factors affecting the image of the scientists. *Eurasia Journal of Mathematics, Science & Technology Education*, *4*, 55–61.
Turner, Y. (1967). *The forest of symbols*. Ithaca, NY: Cornell University.
Van der Veen, J. (2012). Draw your physics homework? Art as a path to understanding in physics teaching. *American Educational Research Journal*, *49*, 356–407.
Van Dijk, E. M. (2014). Understanding the heterogeneous nature of science: A comprehensive notion of PCK for scientific literacy. *Science Education*, *98*, 397–411.
Walls, L. (2012). Third grade African American students' views of the nature of science. *Journal of Research in Science Teaching*, *49*, 1–37.
Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, *50*, 1081–1121.
Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures*. Chicago, IL: Rand McNally.
Wentorf, W., Höffler, T. N., & Parchmann, I. (2015). Schülerkonzepte über das Tätigkeitsspektrum von Naturwissenschaftlerinnen und Naturwissenschaftlern: Vorstellungen, korrespondierende Interessen und Selbstwirksamkeitserwartungen. *Zeitschrift für Didaktik der Naturwissenschaften*, *21*, 207–222. [Students' concepts about scientists' activities: Views, correspondent interests and self-efficacy].
Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität* [Interrater-accordance and interrater-reliability]. Göttingen: Hogrefe.