1. Introduction

Anthropologists and cognitive scientists argued that human specialization in recent

history appears through "reasoning about the causal structure of the world" (Pinker,

2010, p. 8998). The knowledge about the "structure of the world" is increasingly informed

by advances in science, technology, engineering, and mathematics (STEM).

Economists attribute modern societies to be knowledge based, technological, and international

(Friedman, 2005). Economists and developmental pundits unanimously

agree that STEM subjects are amongst the primary factors (if not the prime) to drive

progress (e.g., income growth) and help developing countries to catch up with more

developed countries (Deaton, 2013; IOP, 2012; PCAST, 2012; Pinker, 2018; Radelet,

2015; Rosling & Rosling Ronnlund, 2018; Woodward, Shurkin, & Gordon, 2009). In

order to tackle global problems, e.g., climate change, outsourcing of fossil fuels, or

population growth, the application of STEM knowledge is inevitable (Diamond, 1997;

European Commission, 2014; Pinker, 2018; Sowell, 1997; Wieman & Perkins, 2005).

The importance of STEM is reected by Tate (2001)'s proposal for STEM education

as a civil right, and Snow (1958)'s argument that science knowledge is a "moral

imperative" because of the potentials that STEM o\_ers. The importance of STEM

is reected in estimates that report that future job growth will happen in STEM

(National Academies, n.d.). For example, the European Commission anticipated an

overall 9 percent employment growth for STEM between 2010 and 2020 (European

Commission, 2012).

Consequently, (talented) students in a country need to be engaged in STEM in order

to maintain health, wealth, and prosperity (e.g., Taskinen, 2010). However, the fact

that relatively fewer young people entering tertiary education in the future (changes

in demographics) and the approaching retirement of high numbers of current labor

market participants in STEM are but two reasons for why engaging more students

for STEM is hampered (Dobson, 2014). Estimates state that one million more college

graduates in STEM \_elds in the next decade in the United States are needed to

maintain high living-standards and prosperity (PCAST, 2012). In Germany, in 2016

a peak in vacant job positions in STEM appeared, with over 400; 000 positions vacant

(IW Koln, 2018; ?). However, the interest in STEM subjects in school declines, even for

talented students (summarized in: Taskinen, 2010). In particular, physics is a\_ected in

terms of recruitment and retention of students. For example, neither did recruitment

numbers rose in physics like in other STEM subjects (Evolution of Student Interest

in Science and Technology Studies: Policy Report, 2006; Handelsman et al., 2005) nor

did students retain in college physics (Graham, Frederick, Byars-Winston, Hunter,

& Handelsman, 2013). A still concerning fact is that females compared to males are

largely underrepresented in STEM (Osborne, Simon, & Collins, 2003). For example,

in Schoon (2001)'s large UK sample merely about one in sixty of the girls at the age

of 16 aspired a job in STEM (compared to around one in twenty for the boys). Even

though young women comprise half the workforce in industrialized countries and 58

percent of college-bound population in the United States (Chen, 2013), compared to

young men only 25 percent of the students in introductory physics are females. A mere

20 percent graduating students at bachelor level are female (Hodapp & Hazari, 2015;

Matzdorf & Duchs, 2013; Quaiser-Pohl, 2012), and these numbers decline in the more

inuential positions (Chen, n.d.; Handelsman et al., 2005; Nelson & Brammer, 2010;

Quaiser-Pohl, 2012; E. Smith, 2011).

We report an intervention that seeks to support the physics engagement of young

women and e\_ectively challange the dropout of young women from advanced physics.

The context for this intervention is the physics Olympiad. The physics Olympiad is

a means for identi\_cation and promotion of talented students in physics. The physics

Olympiad is organized in subsequent stages to select a national team comprising a

handfull of the most talented students that compete on an international level against

teams from other countries (for an overview see: Kohler, 2017). The physics Olympiad

appears in adolescence, around 14 to 20 years of age (Kohler, 2017), with the goal

of positively impacting the talented students (Campbell, Wagner, & Walberg, 2000;

Pyryt, 2000; Subotnik & Steiner, 1995). Some reported e\_ects include that students

drawn from the physics Olympiad are likely to pursue physics in their later lives and

succeed in these \_elds (see: Good, Rattan, & Dweck, 2012; Stake & Mares, 2001). In

fact, these students tend to contribute largely to society, as measured through patents,

publications and the like (Campbell et al., 2000). It was also found that cognitive abilities

can be fostered through enrichment programs (Marsh, Chessor, Craven, & Roche,

1995; Wai, Lubinski, Benbow, & Steiger, 2010). There is further evidence that particularly

successful candidates report a positive impact on their future job aspirations in

STEM through such programs (Feng, Campbell, & Verna, 2001; Oswald, Hanisch, &

Hager, 2004; Subotnik, Duschl, & Selmon, 1993). However, an examination of enrollment

patterns indicates that young women have a lower share amongst olympians in

the \_rst stage in the physics Olympiad, and leave the competition disproportionally

over the rounds. For example, in the physics Olympiad around 18 to 27 percent of

participants in the initial round are females, while only 7 percent at the international

\_nal are females (Petersen & Wul\_, 2017; Steegh, Ho\_er, & Parchmann, submitted).

These numbers for the physics Olympiad indicate some underlying problems with

these programs. Assuming that the physics olympians are required to have the highest

quantitative abilities like problem solving in schools (e.g., Heilbronner, 2012; Lubinski

& Benbow, 2006; Phillips, Barrow, & Chandrasekhar, 2002), representative samples

with tests on quantitative abilities at the onset of college predict physics or engineering

would at the lowest comprise 33 percent females (Hyde, Lindberg, Linn, Ellis, &

Williams, 2006). However, considerably less appear in the physics Olympiad. So, why

are fewer young women participating such that many talented students remain an "untapped

source for furthering scienti\_c knowledge" (Ferreira, 2002; Kenway & Gough,

1998). A more balanced participation of young women in the physics Olympiad could

contribute to reverse current trends of public low interest in science and raise overall

support (e.g., \_nancial resources) for these domains (e.g., Hazari, Tai, & Sadler,

2007). The intervention addresses young women in the physics Olympiad, because

these young women are also likely to excel in later physics or STEM careers (Stake &

Mares, 2001), and thus contribute to reduce the problem of disproportionate dropout

for the young women in the physics Olympiad.

Explanations for declining interest and impaired engagement for girls and young

women in physics account for the complex interplay of individual factors with

historical-cultural, structural and social factors (Ceci, Williams, & Barnett, 2009;

Hyde, 2014; Osborne et al., 2003; Scantlebury, 2014). The theoretical framework should

account the various constraints posed upon young women in physics learning contexts

that arise from the environment and from within the young women. Researchers have

argued that the constructs of agency and identity are reective of such mechanisms of

inequality (e.g., Varelas, Tucker-Raymond, & Richards, 2015).

"At their best, [people] are agentic" (Ryan & Deci, 2000, p. 68). Agency was called

the lever in human development (Bandura, Barbaranelli, Caprara, & Pastorelli, 2001).

As such, agency seems to capture a key facet of engagement in a domain. Engagement

in learning settings is a product of individuals' traits and the social context, where

engagement refers to the quality of involvement in academic settings (Bandura, 2018;

Fredricks, Hofkens, Wang, Mortenson, & Scott, 2017). Agency is fostered when the

social context matches with a person's knowledge, values, goals, and interests. Agency

is construed as the "power to originate actions for given purposes" (Bandura et al.,

2001; D. Schwartz, 1997, p. 6). Other authors further this understanding and con\_gure

agency "as a capacity to institute new or unanticipated modes of behavior" (McNay,

2000; Varelas et al., 2015, p. 21).

Agency emerges within social structures (agency-structure-relation). According to

Anthony Giddens, social structures have a "dual" nature. (Giddens, 1981, p. 27) con\_gures

structures as "both the medium and the outcome of the practices which constitute

social systems." Through actions the individual both engages in social structures, but

at the same time actions are the product of surrounding social structures (Giddens,

1979, 1984). Consequently, social-cultural notions of learning stress the importance

of agency, since only through active exploration and intervention an understanding

of the world can be gained (e.g., Gureckis & Markant, 2012), and normative notions

of modern science classrooms endorse the idea that student can exercise agency. For

example, the American Physical Society (APS) endorses an Active Physics Classroom

(Carlone, 2004; Meltzer & Thornton, 2012). The Active Physics Classroom seeks to

overcome traditional instruction and advances ideas such as problems are posed in a

wide variety of contexts, reecting on one's own problem-solving practices, qualitative

reasoning, conceptual thinking, students often work together in small groups, or active

student engagement (Meltzer & Thornton, 2012). Overall, this curriculum is reective

of the requirements for self-directed learning and empowering all students to exercise

agency in the physics classroom.

While agency is a more in-the-moment construct that is malleable on a daily (even

hourly) basis, a student's identity can be conceptualized as a more enduring sense of

sameness and continuity (Erikson, 1950). The development of a personal identity is

the primary goal for adolescent human beings (Erikson, 1968). Developing an identity

implies the exploration of opportunities and the commitment to a set of goals, values,

and beliefs (S. J. Schwartz et al., 2011). An identity comprises parts of the self that

are composed of the meanings that persons attach to the multiple groups and roles

in societies (Stryker & Burke, 2000). Developing an identity is a dialogic process including

the self vis-\_a-vis social structures (C^ote & Levine, 2002; Mead, 1934), where

an individual develops multiple identi\_cations that are relevant for the individual and

a\_ect her actions and choices (Burke, 2003; Stryker & Burke, 2000). In the identity

status model, James Marcia proposed an achieved status where students are committed,

and have an internal, as compared to an external, locus of self-de\_nition (Marcia,

1966, 1980).

In fact, multiple identities attach meaning to the engagement of an individual. In

particular, at least three facets of identity, a personal identity (comprising values,

personality), a social identity (group identity), and a subject identity (a\_liation with

a certain subject) can be di\_erentiated (Burke, 2003). Situational cues that arise in

social contexts can make the identity of a learner salient. Hannover (2000) identi\_ed

situational cues in social learning contexts that are called activation sources that

activate gender-congruent and gender-incongruent self-knowledge (also: Burke & Stets,

1998; Hannover, 2008). These are: the sex-composition of groups (also: Sekaquaptewa

& Thompson, 2003), gender stereotypes, and gender-typical activities. One's gendergroup

("I belong to females/males") becomes an important self-category in many

social contexts (Kessels, 2002). Members of social groups particularly draw information

from social settings "that hold relevance for the value and the status accorded to

their group" (Purdie-Vaughns, Steele, Davies, Ditlmann, & Crosby, 2008, p. 616).

The stereotype inoculation model predicts that particularly in-group experts (e.g.,

female experts as mentors), and the balanced group composition are identity protective

situational cues. Also, social identity theory (Burke & Stets, 1998) and the identity

engagement model (Cohen & Garcia, 2008) predict that the situational activation of

stereotypes exacerbates students' ability to perform and the experience agency.

The subject identity can be conceptualized as student's view of herself as a certain

kind of person (e.g., physics person), and the recognition by others to see this

student as a physics person (Kane, 2012). The following constructs were found to be

integral to physics identity formation for students: competence, performance, interest,

and recognition (Carlone & Johnson, 2007; Cribbs, Hazari, Sonnert, & Sadler, 2015;

Hazari, Sonnert, Sadler, & Shanahan, 2010). More recently, also sense of belonging

was added. Competence entails the students' belief in their ability to be good at the

required tasks and understanding physics. Competence and performance have been

established to be important at the outset of engagement in a domain (Bussey & Bandura,

1999). The belief of self-e\_cacy in a domain, which is similar to competence

and performance beliefs although more task speci\_c, was found to be a variable that

predicted students' performance and later educational outcomes (Pajares & Graham,

1999). It was demonstrated that students on survey items do not distinguish between

competence and performance so that these two are essentially one dimension in empirical

studies (Cass, Hazari, Cribbs, Sadler, & Sonnert, 2011). Interest is the enjoyment

the student has in dealing with physics. Early interest in physics in middle and high

school was found to be a strong predictor of later academic choices (Maltese & Tai,

2011; Tai, Qi Liu, Maltese, & Fan, 2006). For math interest it was shown that it

was related to taking up of advanced courses in math (Koller, Baumert, & Schnabel,

2001). An explanatory link for the relationship between interest and persistence has

been found to be a student's identity in a domain: as they become interested, they

start to see themselves as that "kind of person," and ultimately choose to persist.

Recognition is the students' perception of how much others see her or him as a physics

person. For example, the recognition by meaningful others (parents, peers, teachers)

as a "physics person" is strongly related to having a positive perception of the domain

(Bleeker & Jacobs, 2004; Turner, Steward, & Lepan, 2004). Recognition by others as a

"physics person" in high school has been established to correlate with physics identity

and intended physics career (Hazari et al., 2010). Furthermore, the lack of recognition

can lead to disrupted identities (Carlone & Johnson, 2007). Sense of belonging refers

to positive a\_litation with the domain (Good et al., 2012). Sense of belonging of students

seems to decline in the course of university education (Hausmann, Scho\_eld, &

Woods, 2007). Sense of belonging predicts the intent to pursue math in the future

(Good et al., 2012). The sense of belonging is tied to fundamental human motivations

for closeness with signi\_cant others and interpersonal relationships (Baumeister

& Leary, 1995; Cacioppo, Cacioppo, Capitanio, & Cole, 2015). In sum, these four constructs

facilitate students' subject identity development, and are shaped by the factors

that are outlined in the situated agency model.

We construe a mechanism for inuencing physics identity development to be mediated

through agency that a student experiences in a social context in the domain.

Agency, then, is enabled through an adequate design of the social context that the

student experiences (see Figure 1).

In order to deduce design-principles for an identity-safe social learning context that

supports the physics agency of young women, the intersections of agency and physics

identity with gender will be reviewed. When considering gender-related patterns in

physics, on average, boys and young men, compared to girls and young women, more

readily identify with physics and place a higher personal value on these domains (Eccles,

2011). The reasons for this are explained, at least in part, when considering the

identity constructs of competence, performance, interest, recognition, and sense of

belonging.

With regards to achievement, which relates to competence, the male students in

college tend to outperform females on conceptual understanding assessments (Madsen,

McKagan, & Sayre, 2013). In particular, men have been found to show a greater

variability in many traits such as spatial skills (also aggression) such that in the upper

quantiles men, compared to women, are overrepresented (Hedges & Nowell, 1995).

However, neither are these \_ndings due to di\_erences in innate cognitive ability (Ceci

et al., 2009; Halpern et al., 2007), nor is the representation of men in the upper quantiles

justi\_ed by the variability in the distribution (i.e., there are still disproportionally

more men) (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Hyde et al. (2008) calculated

a proportion of 33 percent females in engineering, where today only around

15 percent are females. Superior competence beliefs of boys are also related to advantages

in physics experiences outside of school. These outside-of-school experiences

can account for perceptions of higher competence as well as higher performance in

physics (Chambers & Andre, 1996; Hazari, Sadler, & Tai, 2008). Young women tend

to perceive physics knowledge in schools as heteronomous and more di\_cult compared

to other subjects Kessels and Hannover (2006), as well as objective and valid for all

time (Driver, Newton, & Osborne, 2000). As indicated earlier, ability stereotypes for

young women in physics learning settings hamper their performance (Kessels, 2002;

Schmader, 2002). Consequently, in the course of their schooling, young women, compared

to young men, acquire depressed beliefs in their ability to perform well in physics

(Fredricks & Eccles, 2002; Kling, Hyde, Showers, & Buswell, 1999).

Boys have been found to express an interest to deal with physics stu\_, both in

school and leisure time, more than girls (Baker & Leary, 2003; Eccles, 1994). Girls

have been found to lose their interest in physics at early ages (Baker & Leary, 2003;

Jones, Howe, & Rua, 2000). Some of the disengagement is accounted for by early

childhood play. For example, boys have more exposure and experience with toys and

objects that align with the traditional physics curriculum than girls do (Ho\_mann,

Krapp, Renninger, & Baumert, 1998; Lubinski & Benbow, 2006; Prediger, 1982). In

particular, a lack of real-world connections and personally relevant content have been

found to result in young women disliking physics (Hazari et al., 2010; Ho\_mann et

al., 1998). Furthermore, the inclination of boys to tinker with physics stu\_ from early

on resulted in a large gap in prior knowledge that bring girls and boys, on average, in

the physics classrooms (Hazari et al., 2008). And, one of the well researched gender

di\_erences in interests predicts that girls, on average, are more interested in people and

communal values, compared to boys, who more endorse things and agentic values (Su,

Rounds, & Armstrong, 2009) that is hunched to relate to di\_erential role-requirements

for women and men in ancient societies (Eagly & Wood, 1999).

Considering recognition, even successful women in science were found to lack support

in their surroundings that would lend to feelings of recognition Carlone and

Johnson (2007). Brickhouse and Potter (2001) highlight that girls in physics class-

rooms often express the feeling of isolation. Similarly, other studies have found that

physics classrooms are often dominated by boys (Labudde, Herzog, Neuenschwander,

Violi, & Gerber, 2000; Stadler, Duit, & Benke, 2000), making it more di\_cult for girls

to perform and be recognized for what they know and can do. Furthermore, the female

gender identity is particularly threatened when it comes to performance situations in

physics (Kessels, 2002). Persistent social stereotypes frame physics as something for

(male) geniuses and brilliant people (Hannover & Kessels, 2004), and high in selfdirectedness

and agency (Diekman, Brown, Johnston, & Clark, 2010; J. L. Smith,

Deemer, Thoman, & Zazworsky, 2014). From as early as 6 years of age these stereotypes

of brilliancy a\_ect the behavior of both girls and boys (Bian, Leslie, & Cimpian,

2017). Stereotypes about women not being as competent in certain domains, such as

physics, can cause depressed performance (Good, Aronson, & Harder, 2008). Physics

environments can even be seen as "hazardous" by the young women who experience

them (Taconis & Kessels, 2009). Such conicts with ones gender identity eventually

lead to disengagement with the domain (Good et al., 2012; Kessels, 2004; Marchand

& Taasoobshirazi, 2013). Such narratives may be particularly problematic for young

women (Leslie, Cimpian, Meyer, & Freeland, 2015), because they limit what female

students believe they can do in the \_eld of physics, in addition to their disinterest when

their goals do not match physics-related careers to begin with (Diekman et al., 2010).

Talented young women in the high school competitions mention a lack of appropriate

role-models for their engagement (O'Connor, 2002). Especially those role-models that

share personality traits in a physics context are important in order for young women

to engage with physics (Taylor, Erwin, Ghose, & Perry-Thornton, 2001) and raise

recognition for the young women. Also the sense of belonging for women in STEM is

impaired (Good et al., 2012; Rattan et al., 2018). Good et al. (2012)'s results indicate

that stereotype threat in math erodes women's sense of belonging to the math community.

In particular, for domain identi\_ed young women stereotype threat impugns

performance and sense of belonging (Good et al., 2008; Schuster & Martiny, 2016). In

the long run the young women turned away from math more likely than their male

counterparts. Similar evidence is given by Brickhouse and Potter (2001, p. 973), who

write that their girls' performance was hampered by the "stereotype threat [...] [of]

being at risk of con\_rming, as a self-characteristic, a negative stereotype about one's

group." Hilts, Part, and Bernacki (2018) found that self-reported relatedness to STEM

was signi\_cantly related to the intent to pursue STEM for young women, but for young

men this link was not signi\_cant. This means that the perceived relatedness/belonging

to STEM is particularly important for young women.

Researchers intervened successfully and empowered young women to pursue physics.

Considerately designed intervention studies that take into account the physics identity

dimensions were able to produce positive results with regards to interest and

engagement for young women. For example, Hau\_ler and Ho\_mann showed that the

adaptation of the curriculum in physics classrooms to the speci\_c interests of girls had

positive e\_ects on physics self-concept, a construct related to physics identity (competence)

(Hau\_ler & Ho\_mann, 2002). Berger (2000) adopted a novel physics curriculum

with contextualized materials (X-ray in medicine) and compared students' interest and

performance to a traditional curriculum. Particularly the young women that lacked

an initial interest were found to bene\_t the most from the new curriculum in terms of

interest and performance gains. Also an active physics classroom (e.g., through handson

experiments) had positive e\_ects on engagement (Jovanovic & Steinbach King,

1998; Palmer, 2009). Even carefully designed small-scale interventions can have positive

e\_ects for students that are historically marginalized in physics learning settings

(Aguilar, Walton, & Wieman, 2014). These interventions account for the potentially

threatening nature of certain social learning settings for particular students' identities

and provide strategies that better protect them from identity threats. With regards to

sense of belonging, Rattan et al. (2018) showed that sense of belonging can be raised

through an appropriate framing of the metatheories about intelligence. They found

that young women who felt that their faculty endorsed universal metatheories (i.e.,

believing that most students have a high scienti\_c aptitude), had a stronger belonging

to the community compared to young women who endorsed non-universal metatheories.

Furthermore, Dasgupta, Scircle, and Hunsinger (2015) showed that female peers

in small-group work can enhance female student's motivation and career aspiration

for engineering. A similar \_ndings was presented by Dennehy and Dasgupta (2017)

where female in-group mentors improve outcomes and retention in engineering. An

important message from these interventions is that even very small-scale interventions

that protect young women from identity threat show abiding e\_ects even after years.

An integrated approach that brings together approaches from multiple interventions

to facilitate engagement for young women in physics settings and promote their physics

identity development may be most bene\_cial. As such, the intervention developed in

this study in the context of the physics Olympiad in Germany capitalizes on many

of these \_ndings. In particular, we will captitalize on the development of sense of

belonging for the participating students since this was motivated to be a particularly

crucial construct for young women's agency and physics identity development.

In order to support young women's agency in physics and development of sense of belonging,

(1) female in-group experts as mentors, (2) the gender group ratio, and (3) the

curriculum and instruction are modelled as potentially threatening mechanisms that

a\_ect young women's agency and sense of belonging in the physics learning context.

(1) Female in-group experts: The "envisioned prototypical student" of a domain

impacts the academic choices of students. Consequently, in-group experts will positively

impact the "envisioned prototypical student" that pursues a domain, and thus

the students' academic choice towards the domain. Therefore female in-group experts

are engaged as mentors in the intervention (Dennehy & Dasgupta, 2017), such that

young women can more readily envision females as prototypical students for physics.

For domain-identi\_ed young women female role models were found to protect performance

(Marx & Roman, 2002), and improve their "implicit STEM self-concepts"

(Stout, Dasgupta, Hunsinger, & McManus, 2011). Furthermore, female in-group mentors

were found to protect sense of belonging for engineering female students (Dennehy

& Dasgupta, 2017). Dasgupta (2011) advanced the idea that in-group experts

play a particularly important role as "social vaccines" to threatened students. Therefore,

in-group experts are means to increase the sense of belonging. In-group experts

implicitly can also lend to the development of self-e\_cacy of young women. As mentioned

above, verbal persuasion in male-dominated domains is particularly important

for women (Zeldin & Pajares, 2000). Thus, the verbal feedback of female role-models is

theorized to be particularly bene\_cial to young women. In general, in-group experts as

role-models, and close mentoring have been emphasized as particularly important for

young women in science (Bartholomew, 1995). More generally, van Tassel-Baska and

Kulieke (1987) stressed the need for talented students to get in contact with practicing

scientists as mentors and role-models.

In this intervention three female experts were engaged to mentor the students in

the intervention (each at one di\_erent location). These female experts were found in

the alumni population of the physics Olympiad. The most successful former female

participants were contacted and all three agreed to participate. All female experts

currently studied physics-related subjects (physics, and mechanical engineering). They

received a training prior to the intervention that lasted approx. 3 hours. In this training

the motivation of this intervention was laid out to them and the learning materials for

the interventions were discussed.

(2) Gender group-ratio: The group constellation in a learning setting was found

to inuence the salience of a student's gender identity (Burke & Stets, 1998; Hannover,

2000). An equal gender-ratio was documented to be bene\_cial (Dasgupta et al., 2015).

For example, when watching a 1:1 ratio conference video women were found to be

protected from psychological distress as compared to more traditional constellations

where men outnumbered women (Murphy, Steele, & Gross, 2007). Similarly, when

young women were in a minority position in a 3-person group their performance in a

math test su\_ered (Inzlicht & Ben-Zeev, 2000). Therefore, in the current intervention

females and males were brought together with a balanced ratio, which would be slightly

unusal for most students. In alignment with the Active Physics Classroom, cooperative

group work was encouraged amongst the students (Meltzer & Thornton, 2012).

(3) Curriculum and instruction: The principles for instruction in the intervention

align with observation in educational research that "people understand concepts

only when they are forced to think them through, to discuss them with others, and

to use them to solve problems." (Pinker, 2018, p. 378). The Active Physics Classroom

encourages the use of real-world problems, inquiry-based learning, and students' interests

and social issues as part of the curriculum{aspects that tend to be neglected in

traditional physics instruction. Students are emphasized to be producers of knowledge

(rather than receivers) in this approach. In the intervention we adopt former materials

that were developed in the context of gender equity in the physics Olympiad (?).

The physics Olympiad focuses on physics problem solving, as one of the central

skills required in physics (Foster, 2000). An e\_ective strategy to physics teaching that

transcends traditional physics instruction (i.e., content oriented) was argued to be

the explicit instruction of central thinking techniques like physics problem solving, or

scienti\_c argumentation (Becerra-Labra, Gras-Mart\_\_, & Torregrosa, 2012; Hu\_man,

1997). Reif (1995) stressed the importance to include instruction on physics problem

solving into the curriculum. Also psychologists argued that making explicit the heuristics

for problem solving bene\_ts the learning process in general (Nickerson, 1994), and

the most apt place to train problem solving is within the domain where the problems

occur (Mayer, 2013). Making thinking explicit causes students to adopt expert-like

behavior more readily such as metacognitive skills (Schoenfeld, 1985). Researchers

recognized the particular bene\_ts to students that identify with historically marginalized

groups (Bogue & Marra, 2007; Fradd & Lee, 1999). For example, Hu\_man (1997)

found in his explicit problem solving the e\_ectiveness of his explicit problem solving

course that particularly supported the young women in his sample (similar results in:

Cooper, Cox, Nammouz, Case, & Stevens, 2008).

In the intervention the students were given a problem solving schema that is adopted

from Heller and Hollabaugh (1992). Furthermore, in order to introduce Newton's laws

the students were given a heuristic of how to tackle such kinds of problems (Reif, 1995).

To enable the learner's mastery experiences with the physics problems, sca\_olding was

provided in the form of formative feedback through the mentors and opportunities for

hints in the solution process. Although Bandura (1986) contended that mastery expe-

rience is the most important source of self-e\_cacy, gender research indicates that additional

aspects needed to be taken into consideration. In particular, women, compared

to men, reported more verbal persuasive experiences than men in their mathematics

learning (Lent, Brown, Cover, & Nijjer, 1996; Lent, Lopez, Brown, & Gore, 1996).

Cooperative group work was utilized to address such important aspects. The verbal

feedback of the female mentors can lend to feelings of self-e\_cacy. Furthermore, the

mentors were instructed to provide positive feedback.

The sample in this study is drawn from students in the physics Olympiad. The students

were invited to participate in this intervention. All participants were young enough

to also participate in the next year's Olympiad. This was important since it was also

tracked whether the students enlisted in the next year's Olympiad as well as an indicator

for physics engagement. The students came together at two subsequent in-person

seminars, each lasting for two consecutive days and half a year apart. Assessment was

pre-post (see Figure 2) with an online survey system (Lime-Survey). Several covariates

and indicator variables (for e\_ectiveness of the seminars) are included that will

be detailed later. The target construct (sense of belonging to physics) was measured

at 5 times (see Figure 2). The hunched mechanisms are displayed in Figure 3.

The host of variables that were utilized in the intervention can be found in the

appendix. As an indicator of reliability, the internal consistencies of the scales, as

measured by cronbach's \_, were calculated. Values over :60 are considered acceptable.

Consequently, sense of belonging (trust) was excluded from the analyses. From the

scale possible science self 3 items were excluded that did not correlate with the whole

scale. 6 items remained in this scale, so that possible science self was still measured

accurately. For the sake of compactness we submitted sample items and the number

of items in each scale in the appendix.

5.2. Recruitment process

Of all invited students 30 percent responded positively and participated in the intervention.

As observed in other longitudinal studies with human subjects, a substantial

dropout of participants was observed throughout the intervention. On average, 54

percent of the participants that initially subscribed to the intervention pursued their

engagement until the end and were included in the analyses as commonly done in

educational assessment (Zhan, Jiao, & Liao, 2017). Therefore, the analyses for e\_ectiveness

of the intervention includes 45 students that persisted until the end (females:

22, males: 23). The dropout will be discussed in the end of the paper in greater detail.

However, the dropout of the males in the control group poses a problematic pattern

with only 31 percent of those students that initially participated persisted until the

end.

In order to evaluate possible deviations between the samples in treatment and control

group versus the general olympian population, a linear model (ANOVA) was used

to \_nd possible e\_ects with respect various covariates, dependent variables, and demographics

at time 1 (see appendix). Relevant di\_erences appear between males and

females for support by peers, and science peer relations, where it was found that males

score higher on these variables.

6. Research questions

The intervention seeks to support young women's physics engagement and physics

identity development in the context of the physics Olympiad. As indicators of engagement

and physics identity we utilized self-reported sense of belonging. Sense of belonging

undergirds the development of a physics identity and was considered a particular

issue of concern for female students. Thus, initial di\_erences and the development of

sense of beloning for the participating female students, also compared to the male students,

are of interest. Relevant covariates (demographics, socio-economic indicators,

social support) were measured and investigated with respect to the time trajectories

for sense of belonging for the participating students. Three research questions structure

the evaluation of the e\_ectiveness of the intervention:

RQ 1: To what extent does sense of belonging develop in the time period where the

intervention occurs for the participating young women and men? To what extent

do covariates impact the development?

As a means for treatment check the students their situational interest (with respect

to tasks, instructors, group instruction). For example, it can be expected that students

who report a high situational interest had agentic experiences and that their physics

identity development compared to students that reported a lower situational interest

was more positive.

RQ 2: To what extent do young women and men report a di\_erent situational interest in

the intervention? To what extent does the aggregated situational interest relates

to the development of sense of belonging for the participating students?

Finally, as a measure of external validity of the intervention, it is assessed to what

extent students that participated in the intervention change their career intention to

study physics and enlist in the following physics Olympiad.

RQ 3: To what extent does the study aspiration for physics of participating students

change? To what extent does sense of belonging predict study aspiration? To

what extent do young women and men that participated in the intervention

enlist in the following competition?

7. Analyses

Modelling a variable (e.g., sense of belonging) over time and accounting for intraindividual

development is considered necessary in order to understand processes of

human development such as identity formation (e.g., Bronfenbrenner, 1979; Cook,

Purdie-Vaughns, Garcia, & Cohen, 2012). Therefore, we capitalize on models that

account for the dependency structure for intra-individual development. In order to

account for the dependency of intra-individual development, linear mixed e\_ects models

have been the method of choice widely applied in social science research (Bryk &

Raudenbush, 1987).

For our speci\_c situation with regards to research design and prior knowledge, many

authors proposed the Bayesian method in order to adequately model the data and for

several reasons:

\_ The worry about sample size is eliminated in the Bayesian method (van de Schoot

et al., 2014). Specifying the model and the priors always leads to inference with

an increasing dependency on priors for small samples.

\_ Bayesian methods yield a straightforward interpretation of results regarding the

probability for our hypotheses of interest (Kaplan, 2014).

\_ The prior information from other studies (Good et al., 2012) can be explicitly

modelled with the Bayesian method. This guards against overinterpretation of

highly unlikely outcomes (van de Schoot et al., 2014).

The Bayesian method places premium importance on predictive accuracy and power

of the proposed models given the data at hand, y. Prior knowledge (in the form of the

parameter distribution, p(\_)) was automatically generated by the "rstanarm" package

(The Stan Development Team, 2018). The generated priors are weakly-informative.

As such, they limit the range of values that are plausible for certain parameters. On

the other hand, weakly-informative priors do not overcon\_dently bias the parameter

estimates towards any particular direction since we have no strong prior knowledge

about such e\_ects (Muth, Oravecz, & Gabry, 2018).

In the Bayesian method the hypotheses of our interest can be evaluated on the

basis of the posterior distribution of the parameters given the data, p(\_jy). We will

test the hypotheses that certain parameters are di\_erent from zero by assessing the

posterior distribution and calculating the ratio of values which are greater than zero.

This property will be denoted with p(\_ > 0), where \_ is the parameter of interest. The

interpretation of this value is that the proportion indicates the amount of con\_dence

with which it can be concluded from this data with these prior values that the given

parameter \_ is greater zero.

In order to assess the quality of the models, Gelman (2004) suggest to perform a

model check based on the posterior distributions of the model parameters. For example,

the distribution of the observed data for sense of belonging can be compared

to the distribution of simulated data based on the posterior values for the parameters.

Furthermore, convergence of the Markov chains was assessed through the ^R-value

of 1 that indicates that the Markov chains mixed properly and converged towards a

stationary distribution. The explanatory and predictive accuracy of the models was

assessed with information criteria that will be discussed later, and a variant of R2 (the

amount of explained variance) in a Bayesian context. This is the proportional reduction

of error for predicting an individual outcome (Gelman & Hill, 2007; ?). In order to

assess the model \_t, \_rst some properties of the posterior distributions in comparison

with the data distribution, p(y), are checked. For example, it is checked how large a

proportion of the models will be greater than the median of the data distribution. The

Leave-one-out statistic (LOO) as a measure for model \_t is calculated. LOO computes

the predictive accuracy of the model on the basis of the log-likelihood, given that one

data point is left out from the modelling process (Vehtari, Gelman, & Gabry, 2017).

8. Results

In order to get an understanding of the interrelatedness of the measured constructs, the

correlations amongst the variables are assessed \_rst. Expectancy of success is signi\_-

cantly correlated with all subscales of sense of belonging (see appendix). Furthermore,

while support by teachers and friends is positively correlated with age, support by

parents is negatively correlated with age. It can be hunched that for older students

parents do not play such an important role for support anymore. Also of interest is the

fact that incremental mindset is signi\_cantly positively correlated with expectancy of

success in physics. Furthermore, socioeconomic background of the parents is of importance.

For example, students with a higher socioeconomic backgroud have a lower

entity mindset and a lower competition achievement. And students with higher socioeconomic

background have a higher study aspiration for physics. Irritating is the

fact that competition achievement is signi\_cantly negatively correlated with sense of

belonging (membership). In fact, all correlations of achievement with the sense of belonging

subscales are negative. From self-concept studies it is known that a higher

performance can be linked to an underestimation (or maybe: more realistic estimation)

of the own competence in a domain (see Kruger & Dunning, 1999). A greater

involvement in the competition and physics community might be a corrective for selfassessment.

The composite measure situational interest (aggregated over subscales and time; and

z-standardized) appears to be signi\_cantly related with the sense of belonging subscales

membership, fading, and acceptance. This means that students who report an overall

higher sense of belonging also reported a higher situational interest in the seminar.

Furthermore, situational interest is negatively correlated with the entity mind-set,

and positively correlated with the incremental mind-set. This means that students

who stronger believe that physics knowledge is not stable and can be trained report a

higher situational interest compared to students who believe that physics knowledge

is stable and cannot be changed. Also study aspiration for physics is signi\_cantly

correlated with situational interest.

8.1. Modelling the development of sense of belonging

RQ 1 considers the development of sense of belonging with regards to predictors such

as gender and group (treatment versus control). It is noted that missings (3 %) were

listwise deleted. The results (see Table 1) indicate that the main e\_ect for time is

not signi\_cant, \_ = 0:01; 95CI = [􀀀0:06; 0:07]; p(\_ > 0) = 0:57. However, the control

group starts o\_ with signi\_cantly lower values in sense of belonging, \_ = 􀀀0:5; 95CI =

[􀀀0:88;0:09]; p(\_ > 0) = 0:01, and there is some evidence that males score higher in

sense of belonging compared to females, \_ = 0:34; 95CI = [􀀀0:03; 0:72]; p(\_ > 0) =

0:96. In fact, there is also some evidence that the control group improved their low sense

of belonging over the course of the intervention, \_ = 0:09; 95CI = [􀀀0:01; 0:18]; p(\_ >

0) = 0:96. Overall, the LOO statistic indicate that the more complex model does not

signi\_cantly improve the \_t to the data and both models explain around 28 percent

of the variance.

RQ 2: Seminar feedback

The next step towards understanding the e\_ects of the intervention is to analyze the

feedback (i.e., situational interest) that the students reported. The means and SDs

(appendix) indicate that all scales are rated above average (5-point scale). At both

times of the in-person seminars, the instructors were rated highest. Almost all students

rated the instructors with the highest possible rating. Overall, the correlations amongst

the scales of situational interest are quiet high, if not always signi\_cant. However, the

subscale instructor does not seem to be signi\_cantly correlated at either time. This is

an artifact of the ceiling e\_ect that is observed for this scale. When all students rated

the instructors literally the same positive then the correlation with other variables is

not meaningful. Since all variables are reasonably highly correlated, and cronbach's

alpha with the aggregated scales and bootstrapped1 95 % con\_dence intervals is 0.78

[0.56;0.91], situational interest will be inserted as a composite scale.2

Regarding gender and group, no signi\_cant di\_erences appear for situational interest

(as tested with an ANOVA with gender and group as predictors), gender e\_ect:

F(1; 41) = 0:14; p = :708; \_2

p = 0, group e\_ect: F(1; 41) = 0:14; p = :706; \_2

p = 0,

gender \_ group e\_ect: F(1; 41) = 1:14; p = :292; \_2

p = 0:03.

In order to examine the impact that the speci\_c design elements, as measured

through the subscales for situational interest (tasks, instructor, and group), had on the

development for sense of belonging, the variability of the slopes (i.e., developmental

trajectories) was estimated. In order to arrive at a parsimonious model a stepwise

exclusion procedure, beginning with a complex model, was conducted (see: Gelman

& Hill, 2007). First, the interaction terms were excluded from the model if they were

non-signi\_cant. Later on, gender and the main e\_ects were excluded if the respective

doubled value of the standard error of the estimate included zero when deduced from

the parameter estimate. Model comparisons were done with ANOVAs (F-tests).

Both models that were considered are very similar in terms of predictive power

(LOO). The di\_erence in LOO is negligible compared to the standard error of LOO.

However, what becomes apparent is that students who report a higher situational

interest have both a higher sense of belonging, \_ = 0:6; 95CI = [0:18; 1:02]; p(\_ > 0) =

1, and a more positive development for sense of belonging compared to students with

a lower situational interest, \_ = 0:15; 95CI = [0:04; 0:26]; p(\_ > 0) = 0:99. This e\_ect

can also be seen in Figure 5. Model 2 explains 23 percent of the variance.

RQ3: What factors predict further participation in the intervention?

Finally, we were interested in changes in study aspiration for physics and the actual

enrollment in the next physics Olympiad (RQ 3). Study aspiration was analyzed with

a regression model where study aspiration (time 5) is the outcome variable. The model

estimates indicate that there is no gender di\_erential e\_ect for study aspiration during

the intervention, neither did young women and men initially di\_er. As done in Good

et al. (2012), a follow-up analysis was conducted in order to check whether sense of

belonging had an impact on study aspiration (an e\_ect that was found in the study of

Good el at., 2012). Table 3 displays the model coe\_cients. Sense of belonging (time

5) turns out to be signi\_cantly correlated with study aspiration while controlling for

initial sense of belonging.

In terms of enrollment in the next physics Olympiad, in the general population,

40 percent (male: 40 %, females: 41 %) of the students enrolled in the next physics

Olympiad. In the treatment group 54 % (male: 58 %, female: 50 %), and in the control

group 47 % (male: 39 %, female: 58 %). A logistic regression was \_t for future participation

(0 =no,1 =yes) in the next competition (Table 4). For this sample it seems that

science peer relations and expectancy of success predict the further participation, but

no other variables appear relevant for prediction. The expected log predictive density

for the full model is greater compared to the restricted model (10:57; SE = 6:1).

9. Discussion

In this intervention the goal was to support the physics engagement for young women

in the context of the physics Olympiad. The spotlight in this paper is on the development

of sense of belonging over the course of a half-year intervention, because the

sense of belonging to the physics community is one central factor that facilitates engagement

and identity development. We sought to identify variables that mediate the

development of sense of belonging. Overall, a six-month intervention was implemented

that o\_ered the young women an identity-safe learning context where they worked on

advanced topics in physics. Previous research on gender di\_erences informed the design

of the intervention. For example, equal gender-ratios, problem-solving instruction, and

female in-group experts as mentors were meant to support the physics engagement of

young women in the sample.

Background checks revealed that the samples initially di\_ered with some respects{a

phenomenon that seems to appear occasionally given that multiple checks with \_nite

samples are performed. Female students report lower science peer relations, and a lower

support by friends and teachers. Findings that can be expected from the literature.

In order to evaluate the e\_ectiveness of the intervention, the following \_ndings are

supported by the data. Considering RQ1 (E\_ects for sense of belonging) it was found

that no main e\_ects for time appeared for the development of sense of belonging. In

contrast to the study by Good et al. (2012) no signi\_cant decrease in sense of belonging

for females over the course of this half year intervention occured. However, important

interaction e\_ects with time and group appeared for sense of belonging subdimensions.

The control group increased from a (relatively) low initial sense of belonging up to the

level of sense of belonging for the treatment group. Furthermore, men, compared to

women, report a higher sense of belonging.

Regarding RQ2 (e\_ects of feedback), it can be said that the situational interest

subscales are rated positive by the students. Especially the subscale instructors had the

highest ratings among young women and men. Situational interest had an impact on

the development of sense of belonging, i.e., the interaction between time and situational

interest is greater than zero. Students that report a higher situational interest are also

the ones who end up with a higher sense of belonging at the end of the intervention.

Since this was only calculated for the treatment group, interpretations that imply

causality cannot be made.

Finally, RQ3 addressed the question of future participation of students in the next

physics Olympiad and their intent to study physics in the future. It turns out that

expectancy of success had a signi\_cant impact on the further participation in the next

physics Olympiad, but no other variables appeared to be related to further enrollment.

However, participants in the intervention (treatment and control) enrolled in the subsequent

physics Olympiad more likely compared to the overall olympian population.

This holds true for males and females, and is regardless of the group (treatment versus

control). Good et al. (2012) presented an analysis of study aspiration in mathematics

and they found that sense of belonging in the post measure was signi\_cantly related

to study aspiration, after controlling for initial sense of belonging. A similar analysis

was conducted in the context of RQ3. The same e\_ect as in the study by Good et

al. (2012) was found: sense of belonging at time 5 was signi\_cantly related to study

aspiration at time 5 after controlling for sense of belonging at time 1. This analysis

indicates that the development of a sense of belonging is a signi\_cant predictor for the

intent to study physics and provides evidence that the increase in sense of belonging

that is observed for some students (particularly those who reported a high situational

interest) is meaningful.

9.1. Limitations

Several aspects are important to keep in mind when interpreting the \_ndings of the

current intervention. First, not all invited participants also participated (33 percent

responded positively). Furthermore, only a fraction of those students that initially

participated persisted until the end (on average 44 percent). Consequently, in the

analyses of the intervention only students that persisted until the end were included.

These students represent a positive selection with respect to motivation to engage

in such an intervention. A dropout pattern appeared either in treatment and control

group of around 50 percent of participants, with the most concerning pattern for young

men in the control group.

A limitation for experimental studies is dropout from the intervention. Dropout of

about 40 to 60 percent for such an intervention program seem common in social research

(e.g., Bijleveld et al., 1998). In order to further understand dropout in this intervention,

it is useful to link drop-out to the covariates. Henceforth, a logistic regression

with dropout as outcome and the covariates as predictors was \_t. In logistic regression

the binary outcome can be predicted with background variables. For dropout in the

treatment group three factors play an important role. First, achievement in the competition

is a crucial factor, b = 􀀀0:23; SE(b) = 0:09;OR = 0:79; z = 􀀀2:56; p < :01. Scoring

one point higher in the competition results in a decrease of the odds of dropout (versus

non-dropout) of 0.79.3 Furthermore, female gender (after controlling for the other

variables) protect students from dropping out of the intervention. Males increased the

odds of dropout by 20.47 (b = 3:02; SE(b) = 1:53;OR = 20:47; z = 1:97; p < :05). Also

situational interest is signi\_cantly related to dropout. Students who report a higher

situational interest decrease their odds for dropout by 0.08 (b = 􀀀2:49; SE(b) =

1:19;OR = 0:08; z = 􀀀2:1; p < :05). No e\_ects were found to be signi\_cant for the

control group.

Another important aspect is the quality of the models. It turns out that the LOO

criteria for the models are only minutely di\_erent from models with no predictor

variables. This indicates that noise is present in the models. Nevertheless we consider

it relevant to report less restrictive models than the only-intercept models since the

found e\_ects are consistent with implications from our theory section (e.g., a high

situational interest and a positive development for sense of belonging are correlated)

and the reader can thus develop a better understanding for potential mechanisms that

are relevant in our intervention.

9.2. Conclusions

Searching for viable interventions that support the physics engagement of talented

young women is an important endeavor in order to raise inclusivity in physics learning

contexts. We know that interventions that target certain variables (e.g., values) can

be e\_ective, and that the adaptation of the social setting can yield bene\_cial results

(e.g., female in-group experts). These insights need to be implemented in the physics

community such that more young women bene\_t from these interventions and that an

understanding for e\_ectiveness outside the laboratory can be established.

We probed for factors (female in-group experts as instructors, gender balanced group

constellation, cooperative learning) that werw found meaningful in order to support

young women's physics engagement. The results point to some important relationships

that sense of belonging has with other outcome measures. On the broadest, it appears

that sense of belonging cannot be impacted in an easy way. A \_ner grained analysis

indicates that the situational interest is a moderating variable and students that are

a\_ected by the intervention do improve their sense of belonging. This means that

students who like the tasks, the group-work, and the instructors bene\_t in sense of

belonging. Sense of belonging is also related to the intent to study physics in the future.

We conclude that sense of belonging is a meaningful construct that can be a\_ected by

a situational adaptation of physics learning environments. However, further research is

needed in order to unravel the mechanisms that are most important to the development

of sense of belonging in physics.

With regards to gender di\_erential e\_ects, we found that the social support by frieds

and the science peer relations appear signi\_cantly higher for young men as compared

to young women. This is an impediment for young women's physics engagement since

peers are the primarly socializing agents in adolescence (Harris & Pinker, 2014), and

are utterly important in puberty (Erikson, 1968). On the other hand, we found no

gender di\_erences for the development of sense of belonging. This gives hope that such

considerately designed interventions can stabilize and foster young women's sense of

belonging and thus bene\_t the physics identity development of young women.