Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #48657

Self-Selection Bias of P-12 Engineering & Computing Activities for Female Pre-College Pupils (Fundamental Research, Diversity)

Michael M. Malschützky, Hochschule Bonn-Rhein-Sieg, University of Applied Sciences, Germany

Michael M. Malschützky is a Research Associate at the Centre for Teaching Development and Innovation (ZIEL) as well as Affiliate Faculty at the Department of Management Sciences at Hochschule Bonn-Rhein-Sieg, University of Applied Sciences (H-BRS), Germany. He received his Diplom-Ingenieur (FH) in Mechanical Engineering from H-BRS in 2005. After working as Test & Validation Engineer (TIER-1) and Program Management Engineer (OEM) in the automotive industry, he returned to academia in 2013, receiving his BSc (2017) and MSc (2023) in Business Psychology from H-BRS.

Gesche Neusel, Bonn-Rhein-Sieg University of Applied Sciences

Gesche Neusel studied Organismic Biology, Evolutionary Biology, and Palaeobiology (OEP-Biology) at the University of Bonn, specializing in Behavioral Biology. She now works at the Gender Equality Office of Hochschule Bonn-Rhein-Sieg, where she is dedicated to promoting and supporting girls in the fields of technology and computer science.

Steven McAlpine, University of Maryland Baltimore County

Steven McAlpine is an assistant teaching professor in the Entrepreneurship, Engineering, and Individualized Study programs at UMBC. He has been teaching role playing game design and leading campus wide RPG events for the past ten years. He also leads the multidisciplinary sustainable design course entitled INDS 430: The Kinetic Sculpture Project, which won the grand prize in the 2015 Baltimore Kinetic Sculpture race.

Dr. Jamie R Gurganus, University of Maryland Baltimore County

Dr. Jamie Gurganus is a Teaching Assistant Professor in the Engineering and Computing Education Program and the Associate Director of STEMed Research in the College of Engineering and Informational Technology. She also directs the Center for the Integration of Research, Teaching and Learning (CIRTL) in the graduate school. Her research is dedicated to addressing the complex challenges of educating and developing engineers, teachers, and communities at every level—from P–12 to post-graduate studies. Her work spans engineering identity and mindsets, global and entrepreneurial competencies, failure culture, first-year experiences in engineering, capstone design thinking, and the integration of service and authentic learning into the classroom and developing future faculty. In addition, she is committed to implementing innovative instructional methodologies and optimizing design through both traditional and non-traditional manufacturing techniques.

Self-Selection Bias of P-12 Engineering & Computing Activities for Female Pre-College Pupils (Fundamental Research, Diversity)

Abstract

Despite targeted interventions, female enrollment in engineering and computer science programs at German universities remains low. To address this disparity, P-12 engineering and computing activities, grounded in Bandura's self-efficacy framework, have been implemented at Hochschule Bonn-Rhein-Sieg, University of Applied Sciences, Germany to empower female pre-college pupils through mastery experiences, vicarious learning, and stereotype reduction. However, longitudinal evaluations of these programs revealed minimal improvements in self-efficacy and STEM-related gender stereotypes, as participants already exhibited high baseline values in these domains prior to engagement. This study investigates whether a self-selection bias exists among participants, hypothesizing that female pupils who voluntarily enroll in such activities differ systematically from the broader population in factors influencing STEM career choices. A comparative analysis was conducted between n = 38 P-12 participants (aged 10–13) and n = 19 demographically age-matched independent female pupils from a local school. Surveys assessed engineering & computing self-efficacy, STEM-related gender stereotypes, general technology interest, school self-efficacy, personal importance of school subjects, and role-model access. Results revealed significant medium-effect differences: P-12 participants demonstrated higher engineering self-efficacy (d = 0.68), weaker gender stereotypes favoring boys in technology (d = 0.66), and greater interest in technology (d = 0.73) compared to the school sample. Additionally, P-12 participants rated their academic competence in and personal importance of school subjects higher, despite similar role-model exposure. These findings confirm a self-selection bias, suggesting that existing recruitment strategies attract pupils already predisposed to STEM. This limits the generalizability of program outcomes and underscores the need for revised outreach to engage underrepresented groups. Practical implications include redesigning recruitment to target pupils with lower self-efficacy and stronger stereotypes.

1. Introduction

1.1. Background and Motivation

The proportion of female students in classical engineering and computer science undergraduate degree programs at German universities remains disproportionately low. For example, at Hochschule Bonn-Rhein-Sieg (H-BRS) the proportion is on average below 12 % and differs only slightly between the offered degree programs Mechanical Engineering (6.8 %), Electrical Engineering (10.8 %), or Computer Science (13.9 %) [1]. An exception to this is Sustainable Engineering with 19.6 % female students [1], a newer degree program with societal references in its designation and associated marketing. This deficiency of female engineering and computer science students in Germany can be attributed to various reasons. Traditional gender- or engineering & computing stereotypes, lacking female role-models in computer science and engineering and low domain related self-efficacy are potential reasons resulting in the observed low rate of female engineering students in Germany.

1.2. Structural factors in the choice of degree program

In Germany, school attendance is free for public schools and compulsory until the age of 18 [2]. Thereby, the German school system is selective, as pupils are separated between different tracks of secondary schools based on their academic performance, but also permeable as the qualification to enroll in a German higher education institutions can be obtained in various ways within the German education system [3]. The range of higher education entrance qualifications (HEEQ) to German universities or universities of applied sciences includes direct tracks by graduating a Gymnasium (academic secondary school) to vocational training (depending on the type of university and its degree program). To address potential overcrowding at universities (of applied sciences) a numerus clausus based on GPA can be applied. However, as competition for places on engineering and computer science degree programs is not as fierce, there are usually no or only low GPA-thresholds.

From a regional perspective, engineering and computer science undergraduate programs are widely available [4], particularly at universities of applied sciences, whose bachelor's and master's degrees are equivalent to those of universities from a legal perspective and in terms of qualification for a PhD-program.

Public universities are tuition-free, only comparatively low social and student body contributions (approx. 350 Euro per semester) are to be paid typically including public transport, student union and student body contributions offering social, economic, and cultural support and participation opportunities. Additionally, governmental student financial aids in the form of loans and grants are offered depending on the parents' income.

In result, the selective school system, tuition fees, regional availability of programs, and numerus clausus do not appear to have a strong negative impact on the choice of degree program.

1.3. Personal factors in the choice of degree program

According to Bandura [5], **self-efficacy** (after appropriate skills and adequate incentives are given) is a major determinant of activity choice, expended effort and persistence against obstacles and aversive experiences. Although developed for research on behavioral change in therapeutic procedures to alter behavior, Bandura's self-efficacy framework is generalizable beyond psychotherapy to other psychological phenomena involving behavioral choices and regulation of effort in activities having adverse effects. In the context of higher education, self-efficacy is therefore a determining factor regarding choice of a specific degree program and how strongly the choice of a degree program is defended against obstacles.

Bandura's framework identifies 4 different sources to develop beneficial efficacy expectations. An enactive source is (1) *performance accomplishments*. It is based on personal mastery experiences in a specific domain. Success raises, repeated failures lower the efficacy expectations, especially if failures occur early in the development process. Once established, a domain specific self-efficacy can get generalized mostly to similar but also to substantially different activities. Less dependable as personal mastery experience and therefore resulting in weaker and more vulnerable self-efficacy is (2) *vicarious experience* by observing others' performance accomplishment. However, observing others, e.g., peers or role-models, performing a task successfully, gives an example of what can be achieved if efforts are intensified and persist without the need to perform the tasks on one's own. With higher similarity with the observed models the beneficial impact rises. Observing various models with differing characteristics succeeding in the task is enhancing the effect on self-efficacy, too. Another mediated source is (3) *verbal persuasion*. Due to its ease and ready availability, it can be used in real tasks, e.g., to mobilize greater effort than without, or without a real task in purely discursive formats. Comparable to similarity with the observed model in vicarious experiences, the persuader's characteristics are an important factor, especially its credibility. The resulting self-efficacy is weaker as in a personal mastery experience of a real task and can be extinguished by disconfirming experience. Especially in therapeutic sessions, the emotive source of focusing on *one's physiological states* (source 4) has an informative function as it can deliver information about personal competency as it is partly used to judge one's anxiety and vulnerability to stress. [5]

1.4. Equal Opportunities Office and P12-activities at H-BRS

1.4.1. Objective & mission statement

To promote gender equality within German universities, *Equal Opportunity Officers (EEOs)* fulfill their obligatory mandate under the German Higher Education Act (e.g., § 24 HG NRW) and underlines the responsibility of universities to actively contribute to their gender equality mandate. They exercise their mandate to implement concrete measures to improve gender equality with their support, participation, and control function across all administrative decisions regarding personnel, social, and organizational measures at the university. Within the equality concept of H-BRS's EEO, P12-acitivities are offered to female pre-college pupils to give them the chance to overcome structural, social, and personal barriers.

The P12-activities at H-BRS are organized and carried out by staff of the Equal Opportunities Office with the aim of empowering female pupils especially in the fields of engineering and computer science and to raise the proportion for female enrollment and retention. Based on Bandura's self-efficacy framework [5], the activities are intentionally designed as corrective experiences to overcome self-debilitating expectations and to give female pupils the opportunity to discover their interest in computer science and engineering through practical experience with like-minded people. Main objective is to establish and enhance a domain specific self-efficacy through the experience of mastery in engineering and computing tasks, which favors the choice of these study programs.

The offered activities range from one-day projects during school hours to one-week courses during the school vacations and are open to female pupils from the 5th grade onwards from all types of German secondary education schools. As successful performance is the primary vehicle of psychological change [5], activities contain different tasks for pupils of grade 5 to 7 and grade 8 and above, to ensure an age-appropriate difficulty level. Too difficult tasks carry the risk of failing, resulting in reducing or even extinguishing prior self-efficacy values. Too easy tasks carry the risk of attributing success incorrectly to external factors, which would also be harmful for the development of self-efficacy.

Important part of the activities' design, offering efficacy-altering experiences, are direct and mediated experience sources for developing a domain specific self-experience according to Bandura's framework (1977):

- To provide an authentic experiential base, tasks cover, for example, soldering, CAD, microcontroller programming, electrical circuits, and game design.
- Vacation activities take place directly on campus in seminar rooms and laboratories of the engineering and computer science departments in order to create as many authentic contextual factors, e.g., social, situational, and temporal circumstances, as possible.
- To support social comparison, the activities are designed mono-educational for female pupils only. Additionally, the group setting gives multiple models in performing the task with a variety of further differing characteristics.
- To reduce a potential attributional error, i.e., attributing success not to one's own capabilities but to external factors, e.g., (task difficulty, fortuitous, or external aids), especially external aids are reduced to a minimum after implementing the needed capabilities needed for the task.

The technical topics and, above all, the announcements of the courses are chosen and formulated in a gender sensitive way that they are attractive to female pupils. Experience shows that interdisciplinary, design, and creative aspects as well as socially relevant issues of engineering and computer science are emphasized.

1.4.2. Example of an offered P12-activity

A vacation course for female pupils in grades 5 to 7, described here as an example, brought together 10 participants for an introduction to CAD. Conducted over four consecutive days from 9:00 AM to 1:00 PM, this program aimed to spark interest in technology and engineering among young female pupils through hands-on learning and practical applications.

The course began with a team-building activity involving the online whisper post game "Gartic Phone", helping the pupils to get to know each other while fostering a sense of community. This activity also served as a playful introduction to digital drawing challenges, highlighting the difficulty of freehand drawing on a computer. Through this exercise, pupils could better appreciate the advantages of CAD technology, particularly in terms of precision and the need for clear instructions when working with software and machines.

To build familiarity with CAD, participants were introduced to two open-source software programs they could continue to use at home. Another program, FiloCAD, is not open-source but developed specifically for educational use and allowed the pupils to create vector files from pixel images. This software, connected to a hot-wire foam cutting machine, provided a simplified entry point without overwhelming users with complex features. The pupils learned the importance of adjusting machine settings for different materials, practicing their skills by creating and decorating foam models to emphasize the creative aspect of CAD design.

Following this, pupils advanced to using Silhouette Studio, a more versatile program that introduced additional design functions. This software phase allowed participants to operate a vinyl cutter, further exploring digital fabrication. They selected either vinyl or heat-transfer foil, creating custom designs for objects or personal clothing items they brought to the workshop.

The final project involved Inkscape, where pupils learned to create files for laser cutting. This task culminated in designing acrylic LED-lit lamps, engraved with their personalized designs and assembled as unique keepsakes. The laser is located in the university's machine hall and workshop, where students also work. The female pupils feel empowered to work with industrial machines and at the same time get to know students as close role-models.

Throughout each session, the students enjoyed daily breakfasts and team-building activities, creating a friendly and relaxed atmosphere that distinguished the course experience from a typical school setting. This focus on collaboration and enjoyment was crucial in making the workshop an engaging and memorable holiday event, motivating students to explore further opportunities in technology and design.

By the course’s end, each participant had gained practical experience in multiple CAD programs and fabrication machines, providing a comprehensive introduction to the digital manufacturing process.

Figure 1 Pictures of P12-activity examples at H-BRS: A) technical drawing with Silhouette Studio with B) resulting folio-cut, C) laser-cut example with CAD file and D) hot-wire foam cutting example

1.4.3. Evaluation of H-BRS's P12-activities

To quantify the impact of the P12-activities, a longitudinal, survey-based evaluation approach (for survey details see section 3. Methodology of this contribution) has been applied since October 2023. With pre-measurements taken directly at the participants' arrival on the first day of a P12-activity and post-measurements as last task on the final day, the P12-acitivities' impact on different variables that have a potential influence on the pupils' decision to enroll in an engineering or computing degree program should be quantified. This approach does not only allow to observe the participants intra-individual development in these factors, but also to compare different activities.

Deviating from the qualitative feedback on the activities from the participants, the sample of n = 85 female participants (M_{age} = 13.52 years, SD_{age} = 1.59) across seven P12-activities at H-BRS in 2023 and 2024, showed only in two of the compared variables (Table 1) a positive development, i.e., computer science related gender-stereotypes (d = 0.18) and general interest in technology (d = 0.15). Beyond that, this positive impact of the P12-activities remained below the threshold of a small effect [6]. The intended positive impact on engineering & computing self-efficacy, a major objective of the offered activities, could not be observed from a statistical perspective.

Table 1
Evaluation of seven P12-activities at H-BRS between 2023 and 2024 on participants personal factors

Variable	**Measurement**				**Comparison**		
	Pre		Post				
	M	*SD*	*M*	*SD*	*t*	*df*	*p*
STEM related gender-stereotypes[a]:							
Mathematics	3.36	0.90	3.42	0.79	0.27	50	.785
Computer Science	2.69	0.92	2.85	0.87	-2.28	50	.027
Technology	2.77	0.87	2.76	0.87	-0.47	50	.642
General interest in technology[a]	3.31	0.94	3.49	0.94	-3.03	50	.004
Engineering & computing self-efficacy[b]	3.99	0.62	3.93	0.76	0.74	50	.465

Note. n = sample size. M = mean. SD = standard deviation. t = t-value of Yuen's t-test for paired samples [7], [8], [9] with 20% trimming [10]. df = degrees of freedom. p = p-value. [a] n = 84. [b] n = 85.

A closer look at the descriptives values offers a possible explanation for the unexpected results. Even prior participation in an offered P12-activity female pupils have high values in engineering and computing self-efficacy (M = 3.99, SD = 0.62, measured on 5-point Likert scale). Moreover, they show, based on one-sample t-tests, only slightly biased stereotypes in favor towards boys regarding computer science ($t(83)$ = -3.09, p = .003, d = 0.34) and technology ($t(83)$ = -2.39, p = .019, d = 0.26) and even a slightly biased stereotype in favor towards girls regarding mathematics ($t(83)$ = 3.64, p < .001, d = 0.40). Although the stereotypes regarding these three domains are statistically significant, the magnitude of difference from a purely unbiased view is considered as small [6]. From this perspective, the female pupils already showed values that were intended to be achieved through the P12-activity before participating in the activity.

2. Research Question

From studies in the field of psychology, it is known that test subjects tend to participate in studies consistent to their needs and characteristic, which can result in biased results due to self-selection [11]. Accordingly, the unexpected observations in the P12-acitivities' evaluations led to the research question (RQ) of this contribution:

RQ. Are the female students who choose to participate in H-BRS's P12-activities subject to a self-selection bias?

In this case their values in variables representing decision factors for choosing a degree program prior their participation would differ significantly from an independent sample of female pupils of the same age range.

3. Methodology

To answer the research question, this contribution relies on a cross-sectional study design utilizing a survey to capture different variables that have a potential influence on the pupils' decision to enroll in an engineering or computing degree program. These variables will be compared between a sample of participants of H-BRS's P12-acitivities and an independent school-sample of female pupils of the same age range.

3.1. Survey description

The survey covers seven variables, including item-batteries for the five psychological constructs (a) *engineering & computing self-efficacy*, (b) *STEM related gender-stereotypes*, (c) *general interest in technology*, (d) *school self-efficacy*, and (e) *personal importance of school subjects*. In addition, the two demographic variables (f) *school grade* the participants visit and (g) *number of male and female STEM role-models* were gathered. Survey instructions and items were formulated in German and in an age-appropriate way according to our target group recruited from middle-school. For better understanding of the items' texts, the term STEM was explained below the survey items as "STEM refers to the subject and professional groups of science, technology, engineering, and mathematics".

To measure the construct of (a) *engineering & computing self-efficacy*, seven items were formulated based on established self-efficacy scales [12], [13], [14]. According to Bandura [5] efficacy expectations are domain specific and measured by formulating domain specific demands. For this, seven different age-appropriate items were formulated in the form of statements covering various aspects of engineering & computing, e.g., "I have the confidence that I can learn a lot about technology.". From a self-report perspective, the participants rate their individual level of agreement with each statement based on a 5-point Likert-type response-scale (1= strongly disagree, 2= disagree, 3= neither agree nor disagree, 4= agree, 5= strongly agree). To determine the interindividual level in engineering & computing self-efficacy for each participant, the individual responses across the seven items are averaged. Based on the response scale, higher numerical values show higher interindividual level in engineering & computing self-efficacy. Three additional distractor-items relating general self-efficacy, e.g., "I have the confidence that I can work in a team.”, were added to the engineering & computing self-efficacy item-battery, to allow additional correlation and reliability analyses.

The construct of (b) *STEM related gender-stereotypes* is measured with 3 items inspired by the 4 item scale by Wolf and Brenning [15]. Instead of using the identical item formulation with a Likert-type response scale once for girls and once for boys, in this contribution a 5-point bi-polar rating scale with the anchors (1= boys to 5= girls) is used. For each of the three subject-matters mathematics, computer science, and technology, the participants were asked to give their personal perspective, if those subjects are typically preferred by girls or boys. Due to the bipolarity of the response scale, numerical values of 3 show no gender bias in the capabilities of men and women, while higher values than the scale mean of 3 show a bias towards women, respectively lower values show a gender bias towards men.

General interest in technology (c) is measured with four self-report items, e.g., “I like reading about technology.”, covering the pupils’ attitudes and behaviors on a 5-point Likert-type response-scale (1= strongly disagree, 2= disagree, 3= neither agree nor disagree, 4= agree, 5= strongly agree). To determine the interindividual level in the general interest in technology, the individual item-responses across the four items are averaged. Based on the response scale, higher numerical scale-mean-values show higher interest.

The variables (d) *school self-efficacy* and (e) *personal importance of school subjects* share the same evaluation strategy using a list of 12 typical school subjects for German middle-schools (i.e., mathematics, physics, computer science, chemistry, biology, German, English, art, music, geography, history, and sports). In case of school domain related self-efficacy, the pupils were asked “What grade do you give yourself in the following subjects?” on a 6-point response scale covering the German pre-university grading system (1= very good to 6= inadequate (failed)). By grading themselves, the students self-attribute a competence level for each of these school subjects representing their school self-efficacy. In case of personal importance of school subjects, the pupils were asked “How important do you think the following school subjects are for your own future?” on a 5-point Likert-type scale (1= unimportant, 2= rather unimportant, 3= (not labeled), 4= rather important, 5= very important). During analyses, the first four school subjects of above were assigned to a subject-group labeled as “*engineering & computing related*”. The other eight subjects were assigned to a subject-group “*others*”. To determine the interindividual level in school self-efficacy and personal importance of school subjects, the individual item-responses across the engineering & computing related and other (not engineering & computing related) subjects were averaged. Based on the response scale, lower numerical mean-values show a higher level of school related self-efficacy, respectively lower personal importance for the subject group. This allows to compare the school self-efficacy and personal importance of the two subject-groups not only between the two samples but also within each sample.

An item with a free text field response for (f) *school grade* the participants visit replaces in the school version of the survey the item with the explicit age of the participant in the P12-activity version of the survey due to privacy reasons. For the comparison of the school- and the P12-

activity-sample an age will be deduced from the school grade based on the typical age range per grade at German schools.

With the single item "Please indicate the number of women and men in your family or closest circle of friends who work in a STEM field.", the pupils were asked to state their *number of male and female STEM role-models* (g) with two free text fields, each one for men and women.

3.2. Sample description

The total sample of this contribution (N = 104) consists of two independent sub-samples, a *school-sample* of n_{School} = 19 female pupils from a comprehensive school and a *P12-activity-sample* of $n_{P12,total}$ = 85 females who participated in seven different P12-actitivites at H-BRS.

The school-sample (n_{School} = 19) was recruited from a comprehensive school in a small suburb of Cologne from 6th (n = 7) and 7th grade (n = 12). From this, an age range of minimum 10 to maximum 13 years can be deduced, although the pupils were not asked directly for their age due to privacy reasons. In general, the school's catchment area shows a limited selection of secondary education schools, as the only alternative is one academic secondary school, which can only be chosen by pupils showing good academic performance. In result, the pupils of this comprehensive school typically show higher variability (with tendency to lower) of academic performance and a variety of social backgrounds.

The total P12-activity-sample ($n_{P12,total}$ = 85) shows an average age of $M_{P12,total}$ = 13.52 years ($SD_{P12,total}$ = 1.59) with a minimum of 10 and a maximum of 18 years across seven activities at H-BRS in 2023 and 2024. In order to obtain a better fit with the school-sample for the planned statistical comparisons, the P12-activity sample was filtered according to the above-mentioned age range from 10 to 13 years. This resulted in a sub-sample of $n_{P12,filtered}$ = 38 in this age range with an average age of $M_{P12,filtered}$ = 12.18 years ($SD_{P12,filtered}$ = 1.14) across six different activities from October 2023 to July 2024.

3.3. Data Analysis strategy

The statistical analyses of this contribution were performed in jamovi [16]. Inferential statistical analyses were utilized conservatively by performing tests two-tailed. Two groups' means were compared using Walrus-module [17] applying robust Yuen's t-tests [7], [8], [9]. For dependent samples trimmed version, for independent samples trimmed and bootstrapped version was applied. A trimming factor of 20 % allows to eschew tests for normality while using 600 bootstrap samples promise estimates with higher reliability [10].

As dataset's missing data were unrelated to the observed and other values in the dataset, all missing data in the dataset could be classified as missing completely at random (MCAR). Accordingly, missing values were excluded from statistical analysis by pairwise deletion. Resulting differences in (sub-)sample sizes are reported for each variable and test.

Originally, the survey was developed and already utilized to evaluate H-BRS's P12-acitivities, especially their impact on the participants (see section 1.4.3 for details). To compare the P12-acitivity-sample with an independent school-sample of the same age range, the P12-acitivity-samples' pre-measurement prior participation will be used as they are unbiased from a potential impact of the visited P12-activities. This allows to evaluate potential differences between the two samples and research potential biases due to the P12-activities participants' self-selection to visit an engineering & computing related activity.

4. Results

Each of the measured variables (a) *engineering & computing self-efficacy*, (b) *STEM related gender-stereotypes*, (c) *general interest in technology*, (d) *STEM role-models*, (e) *school self-efficacy*, and (f) *personal importance of school subjects* will be reported separately for each of the two samples in a first step, prior the samples' values will be compared between the sample-groups in a second analysis-step.

4.1. STEM related gender-stereotypes

Regarding the two subject-matters *mathematics* ($M_{School,math}$ = 3.53, $SD_{School,math}$ = 1.22, 95% $CI_{School,math}$ = [2.94, 4.11]) and *computer science* ($M_{School,CS}$ = 2.95, $SD_{School,CS}$ = 1.31, 95% $CI_{School,CS}$ = [2.32, 3.58]), the school-sample showed descriptive values close the scale-midpoint of 3. This scale-midpoint represents an unbiased view without favor for any gender on the question if girls or boys typically prefer the two subject matters mathematics and computer science. In results, the 95% CIs for the means as well as not-significant one-sample t-tests (math: $t(18) = 1.88$, $p = .076$; CS: $t(18) = -0.17$, $p = .863$) show that the school-sample had a gender-unbiased view on both subject-matters mathematics and computer science.

In these two subject-matters the school sample is not systematically differing (math: $t_{YuenBT} = 1.24$, $p = .170$ and CS: $t_{YuenBT} = 0.77$, $p = .410$) from the P12-activity-sample gathered prior to the activities ($M_{P12,math}$ = 3.32, $SD_{P12,math}$ = 0.91; $M_{P12,CS}$ = 2.76, $SD_{P12,CS}$ = 1.01; n = 37). Figure 2 shows violine-plots for both group comparisons.

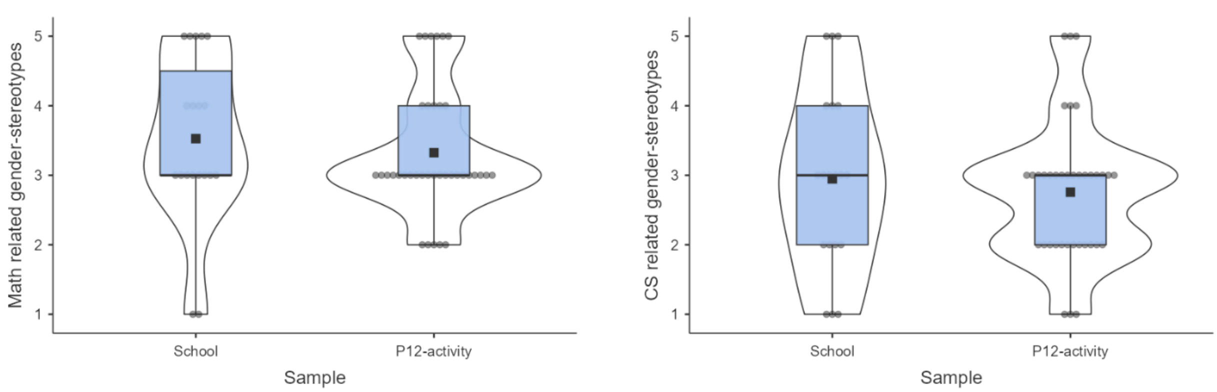

Figure 2. Comparison of school- and P12-activity-sample regarding mathematics and computer science (CS) related gender-stereotypes

Regarding the subject matter *technology* the school-sample ($M_{School,tech}$ = 1.74, $SD_{School,tech}$ = 0.81, 95% $CI_{School,tech}$ = [1.35, 2.13]) showed, based on the 95% CI and a one-sample t-test, a significant gender bias towards boys ($t(18) = -6.83$, $p < .001$). Additionally, the school-sample's values show significant stronger gender bias towards boys compared to the comparable age group of female participants of P12-acitivies ($M_{P12,tech}$ = 2.95, $SD_{P12,tech}$ = 0.91, n = 37) at H-BRS ($t_{YuenBT} = -3.78$, $p = .002$, $\xi = 0.66$, 95% CI(ξ) = [0.41, 0.96]). This difference can be classified as a medium effect [6]. Figure 3 shows violine-plots for the group comparison.

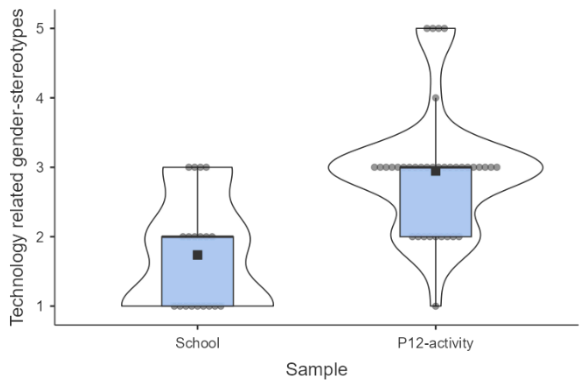

Figure 3. Comparison of school- and P12-activity-sample regarding technology related gender-stereotypes

4.2. STEM role-models

In general, the female pupils of both samples show a small amount of STEM role-models. While the school-sample shows a median number of male role-models of $Md_{School,male} = 2$ and female role models of $Md_{School,female} = 1$, the P12-activity-sample shows comparable medians of $Md_{P12,male} = 1$ for male and $Md_{P12,female} = 1$ for female role-models. Table 2 shows descriptive details for both samples regarding the variable male and female role-models in STEM field.

Table 2
Descriptive values of mixed 2x2-ANOVA across samples as between- and role-model gender as within-factor

Role-model gender	**Sample**	***n***	***M***	***SD***	***Min***	***Max***
Male	school	18	2.50	2.66	0.00	9.00
	P12-activity	29	1.55	1.30	0.00	7.00
Female	school	18	3.22	5.49	0.00	20.00
	P12-activity	30	1.57	1.83	0.00	9.00

Note. *n* = sample size. *M* = mean. *SD* = standard deviation. *Min* = minimum. *Max* = maximum.

A mixed 2x2-ANOVA (Figure 4) with the number of male and female STEM role-models as within-factor and the two samples as between-factor shows that there is neither a significant difference in the number of male and female role-models ($F(1, 45) = 0.90$, $p = .348$), a significant difference of role-models between the two samples ($F(1, 45) = 2.70$, $p = .107$) nor a significant interaction between the gender of role models and the two samples ($F(1, 45) = 0.74$, $p = .394$).

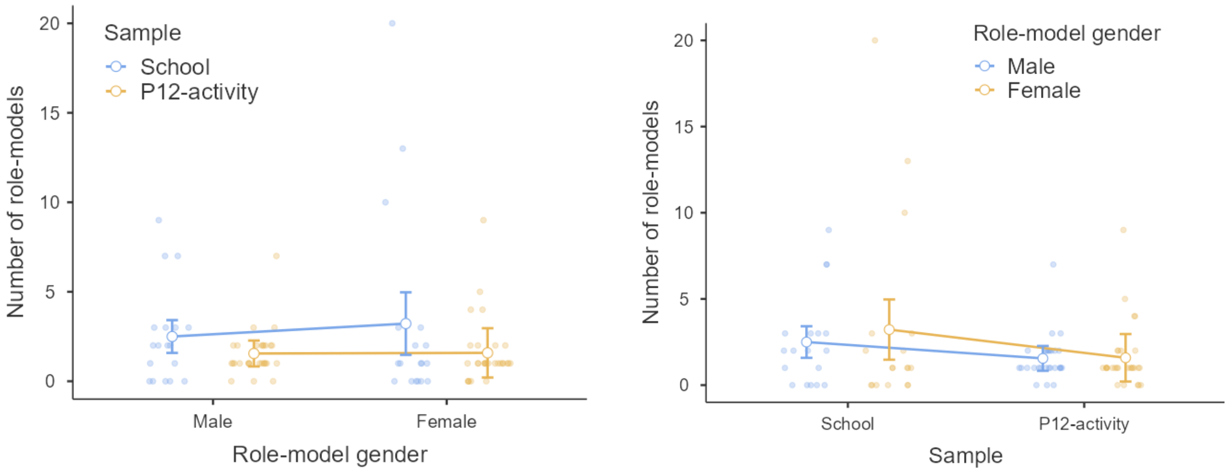

Figure 4. Descriptive interaction plots of estimated marginal means with 95% CI(EMM) of mixed 2x2-ANOVA across samples and role-model gender

4.3. General interest in technology

A reliability analysis of the four items across the total sample ($n = 103$) shows a good internal consistency of McDonald's $\omega = .89$ [18] as well as item-total-correlations [.72, .78] in the preferred range [19]. No improvements can be achieved by eliminating one of the four items.

According to a robust t-test, the school-sample ($M = 2.28$, $SD = 1.03$) showed a significant lower interest in technology ($t_{YuenBT} = -4.41$, $p < .001$, $\xi = 0.73$, 95% CI(ξ) = [0.38, 0.92]) than the P12-activity-sample prior participation in a P12 activity ($M = 3.59$, $SD = 0.93$, $n = 37$). This difference can be classified as a medium effect [6]. Figure 5 visualizes the comparison with violin-plots.

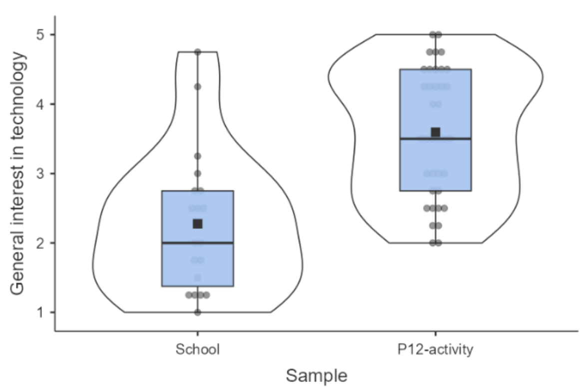

Figure 5. Comparison of school- and P12-activity-sample regarding general interest in technology

4.4. Engineering & computing self-efficacy

A reliability analysis of the seven items across the total sample (n = 104) shows a good internal consistency of McDonald's ω = .85 [18] as well as item-total-correlations [.47, .75] in the preferred range [19]. No improvements can be achieved by eliminating one of the seven items.

According to a robust t-test, the school-sample (M = 3.14, SD = 0.73) showed a significant lower level of engineering & computing self-efficacy (t_{YuenBT} = -4.42, p = .002, ξ = 0.68, 95% CI(ξ) = [0.39, 0.90]) than the P12-activity-sample prior participation in a P12 activity (M = 4.06, SD = 0.72). This difference can be classified as a medium effect [6]. Figure 6 visualizes the comparison with violin-plots.

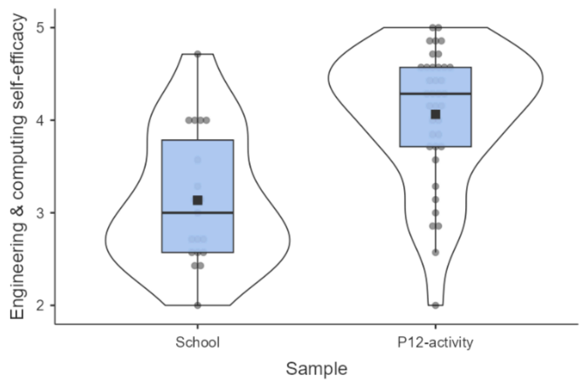

Figure 6. Comparison of school- and P12-activity-sample regarding engineering & computing self-efficacy

4.5. School self-efficacy

To compare the two samples in the construct school self-efficacy, a mixed 2x2-ANOVA with the average self-rated grades of engineering & computing related STEM subjects and other school subjects as two levels of a within-factor and the two samples as between-factor was performed. Table 3 shows descriptive details for both samples regarding the two subject groups, while Figure 7 visualizes the two main effects and their interaction with descriptive interaction plots of estimated marginal means including their 95% confidence-intervals.

Table 3
Descriptive values of average self-rated subject grade for mixed 2x2-ANOVA across samples as between- and school subject-groups as within-factor

Subject-group	**Sample**	***n***	**Average self-rated grade[c]**	
			M	***SD***
Engineering & computing	School	19	2.75	0.77
related[a]	P12-activity	38	1.94	0.61
Others[b]	School	19	2.03	0.52
	P12-activity	38	1.82	0.45

Note. n = sample size. *M* = mean. *SD* = standard deviation. [a] mathematics, physics, computer science, and chemistry. [b] biology, German, English, art, music, geography, history, and sports. [c] German pre-university grading system: 1= very good, 2= good, 3= satisfactory, 4= sufficient, 5= deficient (fail).

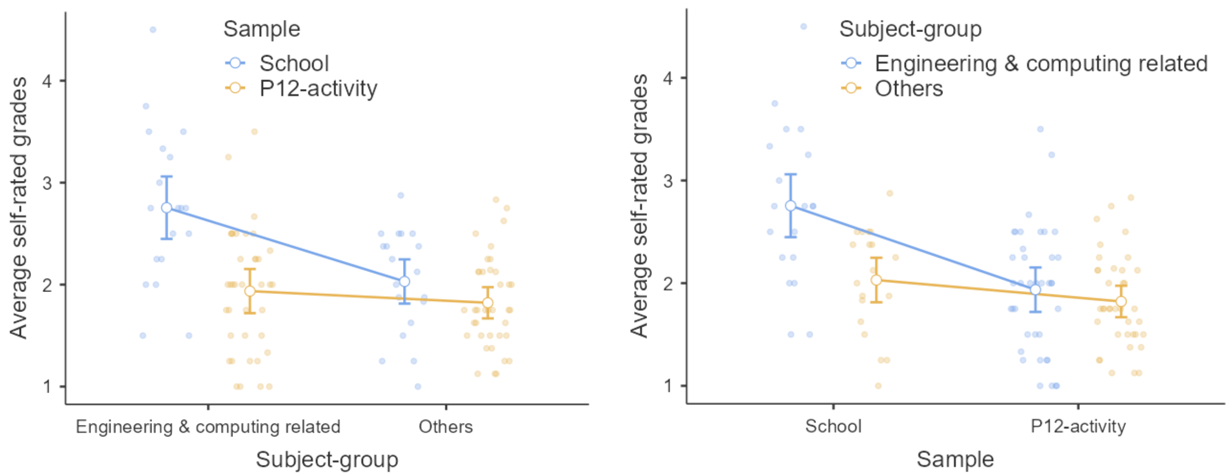

Figure 7. Descriptive interaction plots of estimated marginal means with 95% CI(EMM) of mixed 2x2-ANOVA across samples and school subject-groups regarding average self-rated subject grades

The performed mixed 2x2-ANOVA shows a significant within-effect in the intra-individual competence self-attribution between the subject groups *engineering & computing related STEM subjects* and *others* ($F(1, 55) = 34.47$, $p < .001$, $\eta_p^2 = .39$). In result, the pupils across both samples rate their intra-individual competencies in engineering & computing related STEM school subjects in general significantly lower than in the group of other school subjects. This main-effect can be classified as a large effect [6].

The main-effect between the two samples is also significant ($F(1, 55) = 12.45$, $p < .001$, $\eta_p^2 = .18$). In result, the school-sample rates their competence in school subjects in general significantly lower as the female pupils who plan to participate in a P12-activity. This main-effect can be classified as a large effect [6].

Additionally, Figure 7 shows a significant ordinal interaction between the two subject groups and the two samples ($F(1, 55) = 18.19$, $p < .001$, $\eta_p^2 = .25$). This shows that pupils from the school sample rate their competence in engineering & computing related STEM subjects as lower than female pupils from the P12-acitivity sample, while both samples rate their competence in the other school subjects on a comparable level. This interaction-effect can be classified as a large effect [6].

4.6. Personal importance of school subjects

To compare the two samples, a mixed 2x2-ANOVA with the personal importance of engineering & computing related STEM subjects and other school subjects as two levels of a within-factor and the two samples as between-factor was performed. Table 4 shows descriptive details for both samples regarding the two subject groups, while Figure 8 visualizes the two main effects and their interaction with descriptive interaction plots of estimated marginal means including their 95% confidence-intervals.

Table 4
Descriptive values of personal importance of school subjects for mixed 2x2-ANOVA across samples as between- and school subject-groups as within-factor

Subject-group	Sample	*n*	Personal importance *M*	Personal importance *SD*
Engineering & computing related[a]	school	19	3.01	0.60
	P12-activity	38	3.56	0.86
Others[b]	school	19	3.14	0.68
	P12-activity	38	3.34	0.63

Note. n = sample size. *M* = mean. *SD* = standard deviation. [a] mathematics, physics, computer science, and chemistry. [b] biology, German, English, art, music, geography, history, and sports.

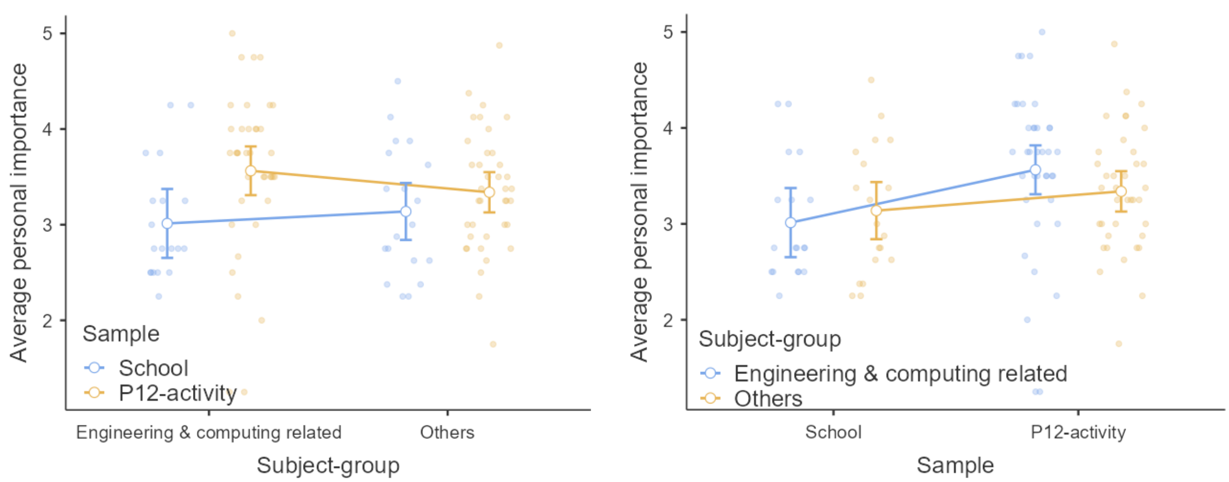

Figure 8. Descriptive interaction plots of estimated marginal means with 95% CI(EMM) of mixed 2x2-ANOVA across samples and school subject-groups regarding personal importance of subjects

The performed mixed 2x2-ANOVA shows no significant within-effect in the intra-individual importance between the subject groups *engineering & computing related STEM subjects* and *others* (F(1, 55) = 0.18, p = .677) as well as no significant interaction between the two subject groups and the two samples (F(1, 55) = 2.15, p = .148).

However, the main-effect between the two samples (F(1, 55) = 5.30, p = .025, η_p^2 = .09) regarding self-reported importance of school subjects in general shows a significant difference between the female pupils in the school-sample and the P12-activity-sample. In result, the school-sample rates their personal importance of school subjects in general significantly lower as the female pupils who plan to participate in a P12-activity. This main-effect can be classified as a medium effect [6].

5. Discussion and Conclusion

Based on unexpected results of the evaluation of P12-activities offered by H-BRS it was hypothesized that the female pupils who plan to participate in these activities are subject to a self-selection bias leading to the observed high values in *engineering and computing self-efficacy* and nearly unbiased *gender stereotypes* regarding the domains mathematics, computer science, and technology already prior participation.

The comparison of the participants' values prior their participation with an independent sample of female pupils from a local school of the same age range showed significant and meaningful differences between the two samples. Compared to the P12-activities-sample, the independent school-sample showed a medium effect of lower *engineering & computing self-efficacy* (ξ = 0.68), medium effect of stronger *gender bias* towards boys regarding the domain technology (ξ = 0.66), and a medium effect of lower *general interest in technology* (ξ = 0.73). Gender-stereotypes regarding the domains mathematics and computer science did not differ. Furthermore, the school sample not only showed a large effect in lower *self-rated competence in school subjects* in general compared to the P12-activities-sample (η_p^2 = .18). The extent of the lower competence self-attribution is even greater in STEM subjects related to engineering & computing compared to other subjects like languages, history, art, or sport (η_p^2 = .25). The same applies to the school-sample's rating of *personal importance of school subjects* in general, which shows a medium effect in lower values compared to the female pupils who plan to participate in a P12-activity (η_p^2 = .09). The *number of female or male STEM role-models* did not differ within each of the two samples nor between them. In summary, these findings indicate a self-selection bias, the P12-activities participants are subject to.

This leads to the practical implication that the self-selection bias must be considered when future P12-activities at H-BRS or their evaluations are planned, e.g. by modifying the recruitment strategies. Otherwise, empirical findings of P12-activities cannot be generalized to those parts of the population that do not have factor values leading to self-selection. Neither will the power of tests be high enough to find significant developments in crucial factors due to only typically small value-changes. The latter makes it difficult to assess the impact of the offered activities and to find the best way to support the target group with thriving opportunities of experiencing mastery to overcome external and especially internal barriers.

Despite the limitation of having a small sample size in the school-sample (n = 19), significant and meaningful differences could be observed indicating a self-selection bias. Especially the small sample size should be addressed in future research to strengthen test-power. Furthermore, samples from different types of schools and age ranges would support external validity of the results.

Concluding, with this contribution we want to invite to a discussion about finding the right recruitment strategies for P12-activities so that a more diverse group of pupils can benefit from the opportunities a P12-activity offers to them as well as empirical findings of such activities can get generalized.

References

[1] Hochschule Bonn-Rhein-Sieg, "StatsInfo 2023 - Hochschule Bonn-Rhein-Sieg." Jun. 27, 2024.

[2] Bundesministerium für Wirtschaft und Klimaschutz, "School system & compulsory education," Make it in Germany. Accessed: Jan. 14, 2025. [Online]. Available: https://www.make-it-in-germany.com/en/living-in-germany/family-life/school-system

[3] "The Educational System in Germany: Case Study Findings," National Institute on Student Achievement, Curriculum, and Assessment; Office of Educational Research and Improvement; U.S. Department of Education, Washington, DC, SAI-1999-3001, 1999.

[4] "Statistische Daten zu Studienangeboten an Hochschulen in Deutschland: Wintersemester 2024/2025," 2024.

[5] A. Bandura, "Self-efficacy: Toward a Unifying Theory of Behavioral Change.," *Psychol. Rev.*, vol. 84, no. 2, pp. 191–215, 1977, doi: 10.1037/0033-295X.84.2.191.

[6] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J: L. Erlbaum Associates, 1988.

[7] P. Mair, R. Wilcox, and I. Patil, *WRS2: A Collection of Robust Statistical Methods*. (Oct. 12, 2022). [R-Package]. Available: https://CRAN.R-project.org/package=WRS2

[8] R. R. Wilcox, *Introduction to robust estimation and hypothesis testing*, 4th edition. Waltham, MA: Elsevier, 2017.

[9] K. K. Yuen, "The Two-Sample Trimmed t for Unequal Population Variances," *Biometrika*, vol. 61, no. 1, p. 165, Apr. 1974, doi: 10.2307/2334299.

[10] R. Wilcox and H. Keselman, "Modem Robust Data Analysis Methods: Measures of Central Tendency," *Psychol. Methods*, vol. 8, pp. 254–74, Oct. 2003, doi: 10.1037/1082-989X.8.3.254.

[11] I. Kaźmierczak, A. Zajenkowska, R. Rogoza, P. K. Jonason, and D. Ścigała, "Self-selection biases in psychological studies: Personality and affective disorders are prevalent among participants," *PLOS ONE*, vol. 18, no. 3, p. e0281046, Mar. 2023, doi: 10.1371/journal.pone.0281046.

[12] AWE-Assessing Women in Engineering, "LAESE - Longitudinal Assessment of Engineering Self-Efficacy," STEM Assessment Tools. Accessed: Dec. 31, 2023. [Online]. Available: http://aweonline.org/efficacy_001.html#desc

[13] M. Jerusalem and L. Satow, "Schulbezogene Selbstwirksamkeitserwartung (WIRKSCHUL)," in *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen*

Begleitung des Modellversuchs Selbstwirksame Schulen, R. Schwarzer and M. Jerusalem, Eds., Berlin, 1999, pp. 15–16.

[14] M. Jerusalem and R. Schwarzer, “Allgemeine Selbstwirksamkeitserwartung (WIRKALL_r),” in *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*, R. Schwarzer and M. Jerusalem, Eds., Berlin, 1999, pp. 13–14.

[15] E. Wolf and S. Brenning, “Wirkung messen: Handbuch zur Evaluation von MINT-Projekten für Schülerinnen.” Hochschule München, 2021.

[16] The jamovi project, *jamovi (Version 2.6.13.0)*. (2024). [Computer Software]. Available: https://www.jamovi.org

[17] J. Love, M. Agosti, and P. Mair, *Walrus: Robust Statistical Methods (Version 2.0.0)*. (2023). [Jamovi module]. Available: https://github.com/jamovi/walrus

[18] R. Hossiep, “Cronbachs Alpha im Dorsch Lexikon der Psychologie,” 2022, Accessed: Jan. 12, 2025. [Online]. Available: https://dorsch.hogrefe.com/stichwort/cronbachs-alpha

[19] A. Kelava and H. Moosbrugger, “Deskriptivstatistische Itemanalyse und Testwertbestimmung,” in *Testtheorie und Fragebogenkonstruktion*, H. Moosbrugger and A. Kelava, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2020, pp. 143–158.