Measuring students' interest in physics

Liza Dopatka

TU Darmstadt, Department of Physics, Physics Education, Hochschulstraße 12, Darmstadt, Germany

Verena Spatz¹, Jan-Philipp Burde², Thomas Wilhelm³, Lana Ivanjek⁴, Martin Hopf⁴
Thomas Schubatzky⁵, and Claudia Haagen-Schützenhöfer⁵

¹*TU Darmstadt, Department of Physics, Physics Education, Hochschulstraße 12, Darmstadt, Germany*

²*University of Tübingen, Physics Education Research Group, Auf der Morgenstelle 14, Tübingen, Germany*

³*Goethe University, Department of Physics Education Research, Max-von-Laue-Str. 1, Frankfurt a. M., Germany*

⁵*University of Vienna, Austrian Educational Competence Centre Physics, Porzellangasse 4, Vienna, Austria*

⁴*University of Graz, Regional Centre for Didactics of Physics, Universitätsplatz 5, Graz, Austria*

Interest is not only regarded a prerequisite for learning motivation but also an educational objective. There is however a major concern regarding the decline during students' school career which has been observed frequently [1]. Therefore, in general education and particularly in physics education great importance is attached to fostering students' interest and a lot of research has been conducted in this regard. To measure students' interest, a number of well-established research scales are available for German-speaking countries, which are primarily based on two constructs of interest. One common construct by Hoffmann, Häußler and Lehrke differentiates between the interest that students show in physical contents and their interest in physics as a school subject [2, 3]. In addition, another construct by Krapp widely used in educational psychology distinguishes between the emotional- and value-related significance of an activity or its content [4]. So far, these two constructs have largely been used unrelated. For the study presented in this paper, they have been connected resulting in four different combinations of emotional and value-related significance attributed to physical contents and physics lessons. These constitute four possible scales which have been integrated in a questionnaire and evaluated in a sample of $N = 1529$ high school students. An exploratory as well as a confirmatory factor analysis has been performed to test the assumed theoretical structure of the underlying constructs. Our data suggests that a revision of the combined scales is necessary as two of the four coincide: all items including emotional valences load on one factor only, while items with value-related valences are distinguishable between interest in physical contents (factor two) and interest in physics lessons (factor three). In this paper we present our study supporting this newly developed measurement model and according research scales for high school students' interest in physics.

I. INTRODUCTION AND MOTIVATION

Modelling and measuring students' interest has always been an important issue in physics education research as demonstrated by the large-scale IPN-, ROSE- or PISA-study [2, 5, 6]. The results of these studies have been widely recognized for several reasons: all studies are based on a very large sample, some reveal an alarming decrease of students' interest over school years - especially in physics - and some suggest contents and contexts to raise and maintain their interest. These studies used different research scales to assess students' interest, derived within several theoretical frameworks using several interest-constructs. This leads to the research question if a combination of these constructs in one questionnaire is possible while still separately representing the different facets of interest. We believe that such combination may lead to a new merged measurement model for students' interest.

This paper introduces the theory of the two interest constructs by Hoffmann et al. [2, 3] and by Krapp [4] before their combination is described and explained. Hereafter, the research methods are presented, resulting in a newly developed measurement model for high school students' interest in physics. This model is then presented along with its statistic quality.

II. THE CONSTRUCT OF INTEREST

Interest is a very multi-faceted term. Here, interest is considered in the context of a learning environment. There are two common constructs that are widely used in the field of physics education research: the person-object-theory [4] and the subject and topic related interest [2, 3].

A. Person-object-theory

Krapp's theory is characterized by the relationship a person establishes with an "object", which can refer to a concrete thing, a topic or a content. Interest-based activities stand out due to emotional and value-related valences at the affective level. Interest is associated with a positive experiential state (e.g. joy) and personal significance or importance. This means that a person's interest can, on the one hand, be influenced by a positive emotional coloring of an object and, on the other hand, by a positive value-related perception of an object [4].

B. Subject and topic related interest

In Germany, early research by the Leibniz Institute for Science and Mathematics Education on predictors of interest in physics classrooms has established the differentiation between subject and topic related interest dimensions. Subject related interest means that the students are interested in the school subject, while topic related interest means the interest in physical topics. This long-term study showed

empirically that these dimensions can be very different and especially that topic-related interest decreases over school years, while embedding topics in adequate contexts represents a promising way to counteract this development [2, 3].

C. Combined interest theory

Table I visualizes the four possible facets of interest, when both constructs are combined. A translated sample item of each scale is listed in table II.

TABLE I. Combined interest theory with four facets of interest.

		interest	
		topic-related	subject-related
valence	emotional	(TE) topic-related interest with emotional valence	(SE) subject-related interest with emotional valence
	value-related	(TV) topic-related interest with value-related valence	(SV) subject-related interest with value-related valence

III. THE QUESTIONNAIRE

The compilation of the written questionnaire is based on the combined interest theory as described above and contains scales for each of the four facets of interest. For this, according scales were identified from literature (scale TE, TV, SV originating from PISA 2006 and the IPN study [2, 7]) or newly designed (scale SE) where no existing scales were available. The following adjustments were made: all Likert scales were adapted to five grading levels from "I do not agree at all" to "I totally agree" and the term "science" was replaced by "physics" wherever necessary. Table II gives a summary of the questionnaire, indicating the number of items, an example of the items, the reliability and the source of each scale.

TABLE II. Scales of the combined construct.

Scale	Number of items	Sample item	Cronbach's α	Source
TE	5	I like to acquire new knowledge in physics.	.92	[7]
SE	5	I like to attend physics lessons.	-	own
TV	5	Physics can explain many processes in nature.	.76	[2]
SV	3	In physics lessons we learn how the content is applied in practice.	.80	[7]

III. METHOD AND SAMPLE

The latent variables are examined using factor analysis: an exploratory factor analysis (EFA) with SPSS as well as a confirmatory factor analysis (CFA) with AMOS by SPSS. The aim of the EFA in a first step is to identify the number of latent variables and the underlying factor structure of the set of variables. This is carried out with two independent samples: Initially, one EFA is conducted with a smaller sample in a preliminary study before another EFA is conducted with a larger sample in the main study also without predetermined factors. In this way, the discovered factor structure in the preliminary study can be confirmed in the main study and two different rotations can be utilized (orthogonal and oblique). In a second step, a CFA with the sample of the main study is used to verify the estimated factor structure and the hypothesized model on indicator level, construct level and model level.

In the preliminary study, data were collected between February and March 2018 with $N = 213$ high school students (49% female, 51% male, 10 classes) of 7th and 8th grade from Darmstadt (Germany) and Vienna (Austria). The data of the main study were collected with $N = 1316$ high school students (52% female, 48% male) of 7th and 8th grade from February to March 2019 (82 classes, 42 institutions). This time the data collection took place in an extended geographical area (Hesse, Bavaria, Styria, Vienna). All of the students attended physics classes, during which the questionnaires were handed out and completed.

IV. EMPIRICAL RESULTS

A. Exploratory factor analysis (EFA)

For the EFA in the preliminary study as well as in the main study a principle factor analysis with Kaiser's criteria for dimensionality (eigenvalue > 1) is applied. Since it is assumed in the preliminary study that the factors are independent, a varimax rotation is used. In the main study the correlation matrix shows that there are relatively high correlations between .30 and .50 for the extracted factors. In this case the rotation is adjusted to a promax rotation. Both EFAs without predetermined factors based on two independent data sets arrive at comparable results: They extract the same three factors according to Kaiser's criteria, which explain 64% or 61% respectively of the total variance. For this reason, only the outcome of the main study, will be presented in more detail below.

The data is suitable for factor analysis as, according to Comrey and Lee, a sample size of more than 1000 is excellent [8]. Moreover, the Kaiser-Meyer-Olkin criterion ($KMO = .95$) and the MSA values (Measure of Sampling Adequacy) are greater than .90 or in one case greater than .80. The communalities of the items are higher than .50. Only items TV_4 (.41) and TV_2 (.29) are below this threshold. Thus, the basic requirements to execute a factor analysis based on the data set of the main study are fulfilled [9].

To assess the suitability of the items, the item difficulties were calculated with the equation of Budischewski and Kriens [10] (minimum value 1, maximum value 5), which are in the range between .36 and .75 and therefore good. Item TE_2 ("I like to read something about physics") has the smallest item difficulty, so only a few learners can agree with the statement. Item TV_1 has the greatest difficulty ("Physics can explain many processes in nature.").

Table III presents the three extracted factors with ordered factor loadings and loadings suppressed below the value .30. It reveals that the hypothesized interest scales (SE, TE, TV, SV) can largely be found with at least three variables loading on each factor. That makes a sufficient base for a meaningful content interpretation [11]. Factor 1 includes all ten variables relating to emotional valence. Therefore, scale SE and TE can be integrated in one scale labeled "emotional valence in physics" (scale EV). Of the items with value-related valence, those concerning the topic, load on factor 2 (scale TV) and, those concerning the subject, load on factor 3 (scale SV).

The selectivity of all items within their factors are above the widely used threshold of .30 [10, 12] and mostly above .50 for good selectivity [12]. This item-total-correlation suggests no item must be removed from a scale. The inter-item correlations are positive, which indicates the one-dimensionality of the items. The values of the mean inter-item correlation (mic) are within the limits of .20 and .40 for scales TV ($mic = .33$) and SV ($mic = .40$) and above this limit for scale EV ($mic = .62$) [13].

TABLE III. Exploratory factor analysis with promax rotation.

Scale	Item	Factor		
		1	2	3
EV = SE + TE	SE_2	.901		
	SE_3	.886		
	SE_4	.872		
	TE_4	.859		
	SE_5	.856		
	TE_1	.842		
	TE_5	.763		
	TE_3	.740		
	TE_2	.680		
	SE_1	.676		
TV	TV_1		.816	
	TV_3		.655	
	TV_5		.508	
	TV_4		.418	
	TV_2		.414	
SV	SV_3			.843
	SV_1			.625
	SV_2			.555

EV: emotional valence in physics, SE: subject-related with emotional valence, TE: topic-related with emotional valence, TV: topic-related with value-related valence, SV: subject-related with value-related valence

TABLE IV. Reliability of the scales.

Scale	Number of items	Cronbach's α preliminary study	Cronbach's α main study
EV	10	.95	.94
TV	5	.74	.71
SV	3	.60	.67

As reported in table IV, the values for Cronbach's alpha of the three produced factors are in a good to very good range. This proves a good reliability on the level of the scales [10].

B. Confirmatory factor analysis (CFA)

The EFAs came to the same results for two completely independent samples. For this reason, the main data set is also used for the CFA. To improve the quality of model fit, the model has been adjusted slightly from EFA to CFA as follows: In the initial model, covariances were only taken into account between the three latent variables. In the improved model, additional covariance relationships between the manifest variables were added by analyzing particularly high modification indices. Thus, the model quality has a χ^2 value of 732.353 and a χ^2/df of 5.859.

To verify this hypothetic interest construct, variances for all three construct variables are set to the value of 1, because thereby factor loadings of all indicator variables can be estimated. The number of parameters is $t = 46$ (18 factor loadings, 10 measurement errors, 10 correlations). With 171 empirical variances and covariances ($p/2 \cdot (p+1)$ with $p = 18$), the number of degrees of freedom is $f = 171 - 46 = 125$. As there is sufficient information available ($t \leq p/2 \cdot (p+1)$), the model can be identified [14].

To select an estimation procedure, the output variables are tested for multivariate normality using AMOS. With the C.R. value of 43.21 for Madia's multivariate kurtosis, this is clearly above the limit of 1.96 [14]. For this reason, Mahalanobis distance values are used to identify and exclude cases from further analysis that differed most significantly. By reducing the sample size from $N = 1316$ to $N = 961$ with this criterion, a C.R. value of Madia's multivariate kurtosis of 1.89 is obtained, which is well below the intended limit. Thus, only a moderate violation of the normal distribution can be assumed, and the maximum likelihood method can be used.

1. Results on indicator and construct level

The results on indicator and construct level are summarized in table V and visualized in the standardized path diagram in Fig. 1. The error variances are calculated with $1 - \lambda^2$ for the standardized estimation results.

Two-thirds of the factor loadings are above .70, suggesting acceptable indicators. The squared factor loadings correspond to indicator reliabilities, where only three indicators (TV_2, TV_1, SV_3) show values below the threshold for good reliability of .40 [15]. However,

eliminating these factors from the model would neither be advisable from a content perspective nor from a statistical perspective, as the number of indicators for the well-established scales TV and SV would become too small. Factor reliabilities (analog to Cronbach's Alpha) are above .70 for all three variables (EV: .96, TV: .77, SV: .73) and demonstrate a high reliability for the results of the standardized solution.

TABLE V. Results of the CFA, reliability calculations and (squared) factor correlations.

	Indicator	Results of CFA		Reliability calculations		AEV ¹
		Factor loading	Error variance	Indicator reliability	Factor reliability	
EV	SE_5	.92	.15	.85	.96	.71
	TE_5	.83	.31	.69		
	TE_4	.85	.27	.73		
	SE_4	.91	.17	.83		
	SE_3	.93	.14	.86		
	TE_3	.75	.44	.56		
	SE_2	.89	.21	.79		
	TE_2	.68	.54	.46		
	SE_1	.81	.35	.65		
	TE_1	.84	.29	.71		
TV	TV_3	.74	.44	.55	.77	.41
	TV_2	.52	.73	.27		
	TV_1	.46	.79	.21		
	TV_4	.66	.56	.44		
	TV_5	.78	.39	.60		
SV	SV_3	.62	.62	.38	.73	.48
	SV_2	.79	.38	.63		
	SV_1	.64	.59	.42		

¹ Average variance extracted per factor

(squared) factor correlations		
	SV	EV
EV	(.46) .68	-
TV	(.79) .89	(.48) .69

FIG. 1. Path diagram of CFA representing the modified model

Validity is checked by looking at the Fornell/Larcker criterion: two latent variables are considered selective, if the squared correlation of two latent variables is smaller than the average variance extracted per factor [14]. This condition is fulfilled for variables EV and SV, the value .46 being smaller than the AEV values of both constructs (.71 and .48). An acceptable separation can also be assumed for the variables EV and TV, the value .48 only slightly exceeding the AEV values of both constructs (.71 and .41). The two latent variables TV and SV correlate very highly (.79) as compared to the AEV values of both constructs (.41 and .48), which is plausible in terms of content. Both constructs deal with value-related valence. Therefore, it seems plausible that statistical selectivity is not quite sufficient. However, these variables have already been found in two independent exploratory factor analyses with different samples, making a strong case for a separation.

2. Results on model level

On this level several statistical tests are used to determine how well a model fits to the data: interference statistics, descriptive statistics and a comparison of the model as derived with EFA and CFA (model 3) with other alternative models: Model 1 consists of one factor and all indicator variables load on one latent variable (41 model parameters). Model 2 consists of two factors, with indicator variables concerning emotional valence loading on the first factor and indicator variables concerning value-related valence loading on the second factor (43 model parameters). Table VI presents the results of the model fit.

Although, according to the chi-square test, criteria of adaptation quality are not quite met by the three models (χ^2/df of 5.859 for model 3 higher than 2.5 [16]), the χ^2 value improves as it decreases from model 1 and 2 to model 3. This may be due to the very large sample size of $N = 961$ students and the strict hypothesis that the empirical variance-

covariance matrix matches the calculated model exactly. The RMSEA value of 0.071, on the other hand, indicates an acceptable model fit for model 3. This result is confirmed by the absolute fit index SRMR = 0.040 for model 3, which is significantly below the cutoff value. Furthermore, all incremental fit indices (CFI, TLI and IFI) for model 3 are above 0.9 and show a good model fit. Model complexity and sample size is taken into account by AIC, BIC and CAIC as well as by SRMR. In comparison with the alternatives, again model 3 shows better values as compared to model 1 and is even slightly better than model 2. On the whole, there are many indices supporting that model 3 is well suited for research in the field of students' interest in physics.

V. CONCLUSIONS

Both methods, EFA and CFA, suggest that a combination of the interest constructs by Krapp [4] and by Hoffmann et al. [2, 3] is possible and that a newly developed model with the respective scales is suitable for measuring high school students' interest in physics. However, when combining both constructs, the EFAs carried out with two independent samples do not extract the four theoretically possible scales, even if they were predetermined. Instead the interest scales of emotional valence and value-related valence in physics could largely be found but subdivide differently: While for the first scale (EV emotional valence) the dimensions of subject- and topic-related interest are not distinguishable, the CFA indicates with slightly better AIC values that it makes sense to separate the value-related valence items according to this criterion into two subscales: TV topic-related interest and SV subject-related interest with value-related valence. In conclusion it can be said that the extracted, new model consists of three scales and has a good model fit with only few structural weaknesses on item level. Therefore, it is suitable for future research, representing three different facets of students' interest in physics simultaneously.

TABLE VI. Results of the model fit.

Criterion	Cutoff-value ¹	Model 1	Model 2	Model 3
<i>Interference statistical quality criteria</i>				
RMSEA	≤ 0.08	0.111	0.074	0.071
<i>(descriptive) absolute fit indices</i>				
χ^2 (df)		1660.214 (130)	801.523 (128)	732.353 (125)
χ^2/df	≤ 2.5	12.771	6.262	5.859
SRMR	≤ 0.10	0.073	0.041	0.040
<i>Incremental fit indices for model comparison (default - independence model)</i>				
CFI	≥ 0.90	0.888	0.951	0.955
TLI	≥ 0.90	0.868	0.941	0.945
IFI	≥ 0.90	0.888	0.951	0.956
<i>Information criteria for model comparison (model 1 - model 2 - model 3)</i>				
AIC		1742.214	887.523	824.353
BIC		1941.801	1096.845	1048.280
CAIC		1982.801	1139.845	1094.280

¹ Thresholds that are commonly used according to literature [15, 16]

-
- [1] P. Potvin and A. Hasni, “Analysis of the Decline in Interest Towards School Science and Technology from Grades 5 Through 11”, *J. Sci. Educ. Technol.* 23, (2014).
- [2] L. Hoffmann, P. Häußler, and M. Lehrke, *Die IPN-Interessenstudie Physik* (IPN, Kiel, 1998).
- [3] P. Häußler and L. Hoffmann, “A Curricular Frame for Physics Education: Development, Comparison with Students’ Interests, and Impact on Students’ Achievement and Self-Concept”, *Science Education* 84, 6 (2000), p. 689-705.
- [4] A. Krapp, “An educational-psychological conceptualization of interest”, *Int. J. Educ. Vocat. Guid.* 7 (2007).
- [5] C. Schreiner and S. Sjøberg, *Sowing the seeds of ROSE. Background, rationale, questionnaire development and data collection for ROSE (The Relevance of Science Education): a comparative study of students’ views of science and science education* (University of Oslo, Oslo, 2004).
- [6] Organisation for Economic Co-operation and Development [OECD], *PISA 2006 Naturwissenschaftliche Kompetenzen für die Welt von morgen Kurzzusammenfassung*, <http://www.oecd.org/pisa/39731064.pdf>. Retrieved 7/05/2020.
- [7] A. Frey et al., *PISA 2006 Skalenhandbuch: Dokumentation der Erhebungsinstrumente* (Waxmann, Münster, 2009).
- [8] A.L. Comrey and H.B. Lee, *A first course in factor analysis* (Erlbaum, Hillsdale/New Jersey, 1992).
- [9] A. Bühl, *SPSS 22. Einführung in die moderne Datenanalyse* (Pearson, Hallbergmoos, 2014), Vol. 14.
- [10] K. Budischewski and G. Kriens, *SPSS für Einsteiger* (Beltz, Weinheim/Base, 2015).
- [11] R.K. Henson and J.K. Roberts, “Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice”, *Educ. Psychol. Meas.* 66, 3 (2006).
- [12] U. Kuckartz, S. Rädiker, T. Ebert, and J. Schehl, *Statistik* (Springer, Wiesbaden, 2013).
- [13] M. Bühner, *Einführung in die Test- und Fragebogenkonstruktion* (Pearson, Hallbergmoos/Germany, 2011).
- [14] K. Backhaus, R. Weiber, and B. Erichson, *Fortgeschrittene Multivariate Analysemethoden* (Springer Gabler, Berlin/Heidelberg, 2015), Vol. 3.
- [15] R. Weiber and D. Mülhhaus, *Strukturgleichungsmodellierung* (Springer Gabler, Berlin/Heidelberg, 2014), Vol. 2.
- [16] C. Homburg and H. Baumgartner, “Beurteilung von Kausalmodellen“, *Marketing ZFP*, 17, 3 (1995).