TPM®

ADAPTATION AND VALIDATION OF THE RATING SCALE DWD-S TO MEASURE DIVERSITY ACCEPTANCE OF ADOLESCENTS IN SCHOOLS

CHRISTOPH JONAS KOLB
UNIVERSITY OF VECHTA, GERMANY

MANUEL PIETZONKA
FOM UNIVERSITY OF APPLIED SCIENCES, HANNOVER, GERMANY

Students can deal with diverse class contexts neutrally, proactively, and constructively or perceive them as irritation or even threat. Diversity acceptance is the individual ability to treat social diversity favorably and acceptably. It captures the individual dealing with diversity globally (across all facets of diversity). This ability is of practical importance when dealing with diverse challenges in schools. The DWD-S (Dealing With Diversity in Schools) scale presented in this paper is a school-related adaptation of Dealing With Diversity in Organizations (DWD-O) scale (Pietzonka & Kolb, 2021), which measures diversity acceptance of adult employees in the working context. The scale adaption took place via three studies: after the item adaptation through interviews with students (Study A), factor analyses (Study B) took place. Through the iterative exploratory factor analyses, 50 models were generated and compared. The CFA confirmed a model with four factors and 13 items. Considering the relevant parameters, the scale has a good model quality. The DWD-S shows good reliability. Construct validity is assessed in Study C. The scale can be used as a research tool to investigate development, changeability, and other implications of diversity acceptance in adolescents aged 13 and older. Practical uses and limitations are discussed.

Keywords: Diversity; Diversity acceptance; Heterogeneity; Adolescence; School; Identity development.

Correspondence concerning this article should be addressed to Christoph J. Kolb, Department of Educational Research, University of Vechta, Driverstraße 22, 49377 Vechta (Germany). Email: christoph.kolb@uni-vechta.de or christophjonaskolb@gmail.com

Teachers and pupils are confronted with an ever-increasing diversity in the classroom context (Clayton, 2010). The causes for the growing diversity include demographic change processes, increasing migration, globalization, pluralization, and individualization tendencies. The influx of asylum seekers and refugees also suggests that subjective and objective diversity in schools will continue to increase (Eurostat, 2022). Living together in diversity is a key issue in today's schools: teenagers and adults with different cultures, ideologies, identities, competencies, and interests must interact and cooperate to achieve common teaching and learning goals. The appreciation of diversity is the basis of an inclusive society and, thus, also of an inclusive educational system. Students may differ in class, school type, moral beliefs, cultural background, religious affiliation, and gender, to name but a few dimensions (Terhart, 2015). Furthermore, learning styles, performance, and learning difficulties also differ greatly in some cases. In addition, there is an increasing diversity due to differences in language competence caused by origin or migration and heterogeneous educational environments. Diverse class contexts impact cooperation, goals, values, working conditions, and the planning, design, and realization of lessons. Diversity can be an asset for a class or working group, as new impulses, ideas, and perspectives can bring benefits; however, the shift toward more diversity also entails risks and uncertainties for the individual in the form of conflicts and misunderstandings. Students may encounter diverse class contexts neutrally, proactively, and constructively or perceive them as an irritation or even a threat (Terhart, 2015).

TPM Vol. 31, No. 1, March 2024 – 105-120 – doi:10.4473/TPM31.1.6 – © 2024 Cises

In school-related diversity research, studies have so far examined how teachers deal with diversity. These are subjective self-ratings of teachers' abilities, attitudes, intentions to discriminate, and behaviors about diversity (e.g., the ATLGB scale; Ensign et al., 2011). There are also instruments to measure the diversity-related school or classroom climate (e.g., the RCI scale; Pike, 2002) and diversity-related discrimination in schools (e.g., the Everyday Discrimination Scale; Clark et al., 2004). There is a need for approaches, however, that focus on the individual abilities of students to deal with diversity. This article attempts to close this research gap. Different constructs that focus on the individual's ability to deal with diversity can be distinguished. These skills "enable […] individuals to deal with social diversity in a constructive and goal-oriented way" (Pietzonka, 2021b, p. 18). Students with this ability work more successfully in diverse classroom contexts. They prove themselves better in various working groups than people perceiving strangeness and otherness as threats or irritation. They perform better than students prone to exclusion, bullying, and discrimination (Walter & Schmidt, 2016). A synopsis of proven measurement methods for the individual level shows that different facets are surveyed: attitudes, normative aspects, conative aspects, knowledge, affective aspects, perceptual aspects, as well as facets of personality dimensions that are expected to be predictors of a professional approach to diversity (Pietzonka, 2018, 2021a). Skills for dealing with diversity can be differentiated in terms of the diversity dimensions considered (Pietzonka, 2016). For example, the dimension of ethnocultural origin is most frequently considered in intercultural competence. In addition, there are also global (i.e., cross-dimensional) approaches to record diversity, for instance, the concept of foreign understanding (Bredella et al., 2000; Bredendiek, 2015), diversity competence (Hoffman & Verdooren, 2019), and diversity acceptance, which is dealt with in this article.

Diversity acceptance captures the individual dealing with social diversity. People with diversity acceptance are more considerate and capable of dealing with diversity in a constructive and goal-oriented way (Pietzonka, 2021b). This definition conceptualizes dispositional ability as broadly as possible. This ability is considered to be of particular practical importance in coping with the diverse challenges in schools since it not only takes individual facets of human diversity into account but also encompasses the individual's handling of diversity globally (across facets). Its bipolar opposite is diversity aversion, the dispositional tendency to react to diversity with dislike and antipathy (Pietzonka, 2021b). As a dispositional ability, diversity acceptance describes differences in feelings and behavior between people. The broad definition explicitly includes the level of consciousness and has social and behavioral components.

It can be assumed that not only dispositional factors but also school and out-of-school socialization and educational processes, as well as individual maturation processes, influence the development of diversity acceptance. An exemplary school-related parameter that could influence individual diversity acceptance is the diversity of students and teachers. Students in more diverse school contexts experience more diversity, which could then influence how they deal with it.

According to social identity theory (Tajfel & Turner, 1979), belonging to social groups helps individuals positioning themselves in their social world. While the connections between social identity development and ethnic diversity are a relatively well-studied area of research (Graham, 2018), the influence of diversity on processes of individual identity development is a research desideratum. Erikson (1968, p. 23) defines identity as a "basic organizing principle" that forms predominantly during adolescence and is characterized by continuity, sameness, and unity. Adolescents must be able to actively participate in social society as part of their identity development (Hurrelmann & Bauer, 2015). Fulfilling these civic development tasks enables them to contribute to social cohesion. In adolescence, significant cognitive, social, and emotional changes occur. In addition to school, other social contexts (such as family, social media, and peers) influence the individual's dealing with diversity. Here, one's identity is negotiated through social interaction,

among other things, with different roles being tested depending on the context. The collective devaluation of minorities can create a sense of community (Tajfel & Turner, 1979). Adolescents can thus value social diversity differently: certain social groups are seen as attractive, and belonging to them represents a resource, while belonging to other social groups is perceived as a flaw or burden. The state of individual identity development could also be related to individual diversity acceptance, so it is expected that students with a diffuse identity status are more likely to be irritated by otherness and less likely to show diversity acceptance than individuals with a fixed identity who are not questioned or threatened by other concepts of life, beliefs, and values about their own identity (Streitmatter & Pate, 1989). If an adolescent is uncertain about his/her identity (identity diffusion), the otherness of their social group can also be an irritation. For example, an adolescent who is uncertain about his or her own sexual identity is more likely to be irritated by homosexuals than adolescents who have a consolidated sexual identity (Timmermanns, 2007).

Overview of Methodological Approach

This paper describes the development and validation of the rating scale DWD-S (an acronym for Dealing With Diversity in Schools) for measuring diversity acceptance of adolescents in the school context in three studies (A-C). This scale is a school-related adaptation of Dealing With Diversity in Organizations (DWD-O) scale (Pietzonka, 2021b), which measures diversity acceptance of adult employees in the working context. The scale makes it possible to examine the students' individual dealing with social diversity and thus closes an important research desideratum. The adult scale is available in English (Pietzonka & Kolb, 2021) and German (Pietzonka, 2021b). The items of the DWD scales contain statements to which the respondents are to assess the degree of agreement or disagreement using a discretely graded response scale. The respondents are first asked: "To what extent do you agree with the following statements?" When calculating the scale value, the individual answers per item are considered as follows: 0 = *completely disagree*, 1 = *mostly disagree*, 2 = *slightly disagree*, 3 = *slightly agree*, 4 = *mostly agree*, and 5 = *completely agree*. The German version of the scale was translated into English by three independent experts and translated back by two native speakers. The back-translation did not reveal any significant deviations from the original. In addition, the quality of the translation was tested in advance by bilingual adolescents (N = 179) who completed both the German and English items. The correlations of the item pairs were all large and highly significant.

In Study A, items of the DWD-O were first adapted to the school context in terms of content and then modified in terms of language so that they are understandable for students aged 13 and older. The age appropriateness was tested through an interview survey. The results were the basis for the item adaptation. Subsequently, the new items were tested within the framework of EFA and CFA about their dimensionalization in Study B. In Study B, paper-pencil questionnaires were made available to all schools or classes that had agreed to participate in the project (nonprobabilistic sample). With this sample, item analyses were conducted, followed by exploratory factor analyses (EFAs) and confirmatory factor analyses (CFAs). The new scale for the school context was tested concerning its construct validity and reliability in Study C. The objectives and methodological procedures of the three studies are described in Table 1. All data collection was carried out in German schools. All relevant German school types were considered: Hauptschule, Realschule, Gymnasium (secondary school), Fachgymnasium, Gesamtschule (comprehensive school), and Privatschule (private school). The evaluated data matrices (in SPSS and in Excel incl. Coding System) as well as the results of the relevant studies (frequency distributions, descriptive statistics, results of the EFA

and CFA for the final model as well as the analyses for reliability and validation of the scale) can be found under the following link: http://doi.org/10.6084/m9.figshare.23244095

TABLE 1
Overview of the data collection of Studies A-C

Data collection	Function of the sample	Participants	Gender and age
Qualitative interview (2017)	Study A: Testing of age appropriateness and adaption of the items	N = 20 from secondary school (10-14 years) (divided into two groups of 10 students each)	50% male, 50% female, M_{age} = 12.9 years
Total Sample 1 (2018)	Study B: Questionnaire survey for item, factor and reliability analyses and construct validation	N = 1031 students from all school types (13-19 years), 995 of them completed the questionnaire; N = 717 after outlier analysis (final total Sample 1)	55.4% female, 42.1% male (missing 2.5%), M_{age} = 15.2 years (SD = 1.5 years)
Subsample 1a	For the EFA	n = 417 students	
Subsample 1b	For the CFA	n = 300 students	
Sample 2 (2020) (panel)	Study C: Questionnaire survey retest reliability, further validation	N = 134 students from secondary school (n = 34 of the 134 for the retest)	53% female (n = 18), 47% male (n = 16), M_{age} = 14.1 years (SD = 0.70 years)

Note. Samples 1 and 2 represent surveys of students and are classified as nonprobabilistic. EFA = exploratory factor analysis; CFA = confirmatory factor analysis.

Linguistic Adaptation of the Items through Qualitative Interviews (Study A)

For the scale adaptation, items that are linguistically adapted to young people's everyday life were needed. The content and metaphors used in the items must also be understandable for adolescents (Kiegelmann, 2009). Diersch and Walther (2010) point to a strong context effect in information about personal attitudes when interviewing children and adolescents, which always presupposes knowledge and evaluation. For this, it is indispensable that the young people surveyed understand the implications of the items. If they can paraphrase them in their own words, it can be assumed that a general understanding prevails.

In order to investigate the age appropriateness of the items, individual interviews were conducted on the items of the preliminary version of the adult scale with a case selection of N = 20 students aged 10-14 years, who were selected randomly using drawing numbers. Beforehand, the items were adapted to the school context. The selection of the students for the voluntary survey was made by drawing lots in two different school classes. The first group of students (n = 10) was asked to read the items aloud and reproduce the respective content in their own words to check how much was understood. If items could not be paraphrased clearly, follow-up questions were asked to clarify comprehensibility. The interviews were recorded, transcribed, and analyzed. It quickly became apparent that some items were not understood correctly. The results show that the items of the adult scale, which had been adapted to the school context, represented a great challenge in terms of their wording and the complexity associated with them. It turned out that several respondents needed help to reproduce the content of certain items in their own words. Concepts behind minorities, morality, or values, for example, were complex for them to grasp.

The misunderstandings, follow-up questions, and hints were documented, transcribed, and evaluated. These misunderstood items were identified in order to simplify them. Complex terms that require a high level of abstraction were not used anymore. The modifications did not change the content implications of the items. Furthermore, we attempted to remain as close as possible to the original in terms of language. After adapting the items, the procedure was repeated with the second group of students (n = 10) to check the linguistic adaptation of the items. The adapted items based on this evaluation were tested with the second half of the students. The intentions of the adapted items could largely be paraphrased by students aged 13 and above in their own words, so the scale is expected to be understandable for young people aged 13 and above. Younger respondents still needed help understanding and could not clearly explain the respective intentions in their own words.

Development of the DWD-S Scale through Item and Factor Analyses (Study B)

After the linguistic adaptation of the items, a survey based on the remaining 25 items of the previous version of the DWD-S was drawn for item and scale testing (N = 1031 participants). For this purpose, a quantitative questionnaire survey was conducted with students aged 13-19 years. When selecting schools and classes, a representative sample was attempted to represent the population. This goal could not be realized because numerous schools refused to participate or canceled at short notice. Within the framework of the questionnaire survey, various sociodemographic parameters were collected: age, gender, cultural background, local background, federal state, grade level, educational background of parents, school type, and intended educational attainment. Within the framework of the questionnaire survey, other scales were collected for validation purposes. Questionnaires with missing values in the DWD-S scale were eliminated. Additionally, the time it took to complete the questionnaire was recorded in three classes (n = 83). The time needed to complete the entire questionnaire was between 11.5 and 32.0 minutes (M = 20 minutes, SD = 4.5 minutes). Incomplete questionnaires were excluded, so 995 questionnaires remained. Due to the questionnaire length of 111 items and the associated stress on the students, a multivariate outlier analysis was useful for detecting anomalies in the data. Furthermore, eliminating outliers is an essential prerequisite for the subsequent inferential statistical procedures. Reactance, fatigue effects, and concentration problems are assumed to be causes of outliers (Moosbrugger & Kelava, 2007). The method used was a projection pursuit approach (PP; Pena & Prieto, 2001). This is preferable to a purely multivariate method such as the robust Mahalanobis distance (Filzmoser & Hron, 2008) because the covariance structure is not initially estimated from the data, thus avoiding singularity problems. The outlier analysis identified 278 anomalies in the first sample, leaving 717 cases for the factor analyses.

The following description of total Sample 1 refers to the remaining 717 cases after the outlier analysis. The average age of the respondents is 15.2 years (SD = 1.5 years). Furthermore, 55.4% (n = 397) of the respondents are female, 42.1% male (n = 302); no answer 2.5% (n = 18). Regarding their origin, 22.3% (n = 160) of the respondents come from a village background (≤ 2,000 inhabitants), 24.4% (n = 175) from a small-town background (2,001-20,000 inhabitants), 23.2% (n = 166) from a city with a population between 20,001 and 100,000 inhabitants, and 22.8% (n = 163) of the respondents come from a large city (> 100,000 inhabitants); no information was provided on n = 53 (7.4%).

Item analysis was used to assess the performance of the items regarding item selectivity and item difficulty. After eliminating outliers and the item analysis, the total sample was divided into two parts using quota-based random numbers so that two separate subsamples were available for the subsequent factor analyses: Subsample 1a (n = 417) was used for the EFA; Subsample 1b (n = 300) was used for the CFA. Before the CFAs, the prerequisites for implementation were checked, all of which were fulfilled:

- interval-scaled variables,
- approximate normal distribution of the variables,
- normal distribution (Shapiro-Wilk test),
- it was checked whether substantial correlations existed in the data set so that factor analyses could be carried out.

The subsequent EFA was iterative, that is, several factor analyses were carried out one after the other. For this purpose, the five cut-off values, .30, .35, .40, .45, and .50, were tested with the R software. The items whose highest factor loading was smaller than the cut-off value tested were eliminated. The item with the lowest total maximum factor loading was then excluded. Thus, 10 individual factor analyses were conducted for each cut-off value (from a single factorial model to a 10-factor model) so that 50 models could be generated and compared. All combinations of the cut-off value and the number of factors to be extracted were analyzed in R with the following algorithm (Pietzonka, 2021b, p. 20):

- Step 1: choose a cut-off value and several factors to extract,
- Step 2: perform an EFA,
- Step 3: the algorithm ends if all maximum factor loadings are ≥ the cut-off value. If there is at least one maximum factor loading < cut-off value, go to Step 4,
- Step 4: remove the item with the smallest maximum factor loading. Then go to Step 2.

After each item exclusion, a new analysis was carried out without the eliminated item. The respective factor analysis was completed as soon as all items with a factor loading ≥ the cut-off value loaded on at least one factor. Varimax rotation procedure in R (Lavaan package Version 0.6-1) was used here. A principal component analysis (PCA) was rejected: PCA would rotate the coordinate axes orthogonally, whereas factor analysis allows arbitrary angles between the axes. The following four criteria were considered in the analysis of the models:

- interpretability of the factors in terms of content,
- level of variance elucidation (the more, the better),
- number of items involved (the fewer, the better), and the
- *p*-value of the χ^2 fit test ($p > .05$).

Subsequently, a CFA was carried out with Subsample 1b ($n = 300$) to test the models that have proven themselves in the EFA within the framework of the four criteria mentioned above. The estimation was based on interval-scaled characteristics using the maximum likelihood method. After the χ^2 fit test, the following parameters were collected to assess the quality of the model and interpreted according to the criteria of Hu and Bentler (1998): root-mean-square error of approximation (RMSEA), standardized root-mean-square residual (SRMR), comparative fit index (CFI), and the Tucker-Lewis index (TLI). Based on these four criteria above, the three models A-C were selected for further analysis. Within the framework of the CFA with Subsample 1b ($n = 300$), the relevant characteristics and parameters for the model quality were analyzed for the models A-C in comparison (Table 2).

All models A-D show a significant result of the χ^2 test in the CFA, which is not unusual due to the large sample and the number of items. Considering all relevant parameters, Model D with $k = 4$ factors and 13 items and a cut-off value of .35 appears adequate and acceptable overall: the RMSEA is .047, 95% confidence interval [.068, .086]. The value of the SRMR is .047, the TLI is .849, and the CFI is .880, which according to Hu and Bentler (1998), all indicates good model quality. The proportion of explained variance is 47%. Model D is thus the final model of the DWD-S scale. Table 3 shows the factor loadings of the DWD-S. The items and the 6-point rating scale of the DWD-S are presented in Table 4.

TABLE 2
Characteristics of models A-D and quality of model fit in comparison

	Model A three factors	Model B four factors	Model C five factors	Model D four factors
Cut-off value	.45	.40	.40	.35
Cumulative proportion of explained variance	.42	.43	.53	.47
Number of items	13	16	14	13
p value of the χ^2 test	.00**	.00**	.00**	.00**
Satorra-Bentler χ^2	514.43**	443.03**	223.94**	311.12**
SRMR	.130	.118	.035	.047
RMSEA	.144	.134	.048	.077
CFI	.769	.810	.959	.880
TLI	.740	.792	.941	.849

Note. Based on Subsample 1b, $n = 300$. SRMR = standardized root-mean-square residual; RMSEA = root-mean-square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index.
** $p < .01$.

TABLE 3
Factor loadings of the DWD-S (Model D)

Item	F1	F2	F3	F4
5		.64		
2		.67		
1		.61		
13	.73			
4	.76			
7			.53	
10			.50	
6				.63
8	.69			
9	.53			
12				.50
3				.36
11			.47	
SS loadings	2.88	1.60	1.36	1.36
Proportion variance	.19	.10	.09	.09
RPV (%)	40.40	21.28	19.15	19.15
Cumul. var.	.19	.29	.38	0.47

Note. F = factor, based on Subsample 1a, $n = 417$; cut-off value: .35; SS loadings = sum of the squares of the factor loadings; RPV = relative proportion variance = proportion variance/.47; Cumul. var. = cumulative proportion of variance explained.

TABLE 4
Items of the DWD-S

	Item	Factor
1	I recognize quickly if someone is treated unfairly	2
2	When I realize that someone is being bullied, I tend to look away*	2
3	In our school it should be even more important to deal with minorities (for example, gays and lesbians or people with a different skin color) respectfully	4
4	I treat foreigners without prejudice	1
5	I am good at integrating outsiders into my group	2
6	We should pay less attention to the needs of minorities (for example, disabled people or people with a different religion)*	4
7	Attitudes of others that differ from my own are easy for me to bear	3
8	I treat gays and lesbians without prejudice	1
9	I treat disabled people without prejudice	1
10	People who are different from me give me a funny feeling*	3
11	I only enjoy working with classmates when they share my ideas and beliefs*	3
12	The more similar students in a group, the more successful the collaboration*	4
13	Toward people with a different religion, I act without prejudice	1

Note. DWD-S = Dealing With Diversity in Schools. The respondents are first asked: "To what extent do you agree with the following statements?" When calculating the scale value, the individual answers per item are considered as follows: 0 = *completely disagree*, 1 = *mostly disagree*, 2 = *slightly disagree*, 3 = *slightly agree*, 4 = *mostly agree*, and 5 = *completely agree*. Items with an asterisk (*) are inversely formulated.

In determining the order of the items, an attempt was made to avoid order effects (Tourangeau & Rasinski, 1988). Missing values are not possible since all items must be answered. It must be considered that the five items with an asterisk (*) are inversely formulated and are included inversely in the calculation. For each factor (F1 to F4), a value can be calculated by taking the arithmetic mean of the corresponding response scores. In addition, a total value is to be calculated from the factors. The relative proportions of the explained variance per factor are obtained by forming the quotient "PV = proportion variance/.47," which results from Table 3. The total value is calculated by the following formula, which takes the mean values of the four factors into account, weighted according to their variance explanation:

DWD-S = .404*F1 + .213*F2 + .192*F3 + .192* F4.

Or in long form:

DWD-S = .404*[(Item 4 + Item 8 + Item 9 + Item 13)/4] + .213*[(Item 1 + Item 2 + Item 5)/3] + .192*[(Item 7 + Item 10 + Item 11)/3] + .192*[(Items 3+Item 6+Item 12)/3].

Table 5 contains the item characteristics and the assignment of items to the four factors. The items of the first factor, "stereotyping", address the degree of bias against minorities (Items 4, 8, 9, and 13). The second factor, "dealing with discrimination" (Items 1, 2, and 5), includes conative aspects that allow the success of specific actions to be assessed subjectively. Factor 3, "affective diversity components" (Items 7, 10, and 11), takes three items that capture emotions when cooperating in diverse social situations into account. They assess how it feels for the respondent to deal with social diversity. Factor

4, “diversity beliefs” (Items 3, 6, and 12), measures the extent to which diversity is seen as valuable for cooperation. It also deals with items with a normative character and address the societal treatment of minorities. There are no correlations between the DWD-S and the school type of the students. Positive correlations can be found between the DWD-S total score and the grade level (.12**) and the desired educational attainment (.25**), whereby the desired educational attainment primarily explains the correlation with the grade level. A gender-specific difference is significant in the Mann-Whitney U-test (z = –5.77, p = .00): female respondents show more diversity acceptance than male respondents.

TABLE 5
Item and scale characteristics

Item	Factor	*SD*	P_i
1	2	1.14	62.5
2	2	1.28	54.7
3	4	1.07	67.7
4	1	1.29	65,5
5	2	1.24	51.7
6	4	1.10	63.8
7	3	1.13	58.7
8	1	1.24	71.0
9	1	1.19	69.3
10	3	1.20	57.8
11	3	1.28	60.2
12	4	1.64	48.2
13	1	1.23	65.7

Note. P_i = item difficulty, based on total Sample 1, N = 717.

Table 6 shows the intercorrelations of the subscales with each other. All correlations have a p < .001 and are statistically significant, considering the Bonferroni-Holm correction. In total Sample 1, the DWD-S total score is M = 3.81 (SD = 0.65).

TABLE 6
Intercorrelations of the subscales

	F1	F2	F3	F4	*M*	*SD*
F1	1				4.11	0.93
F2	.50**	1			3.39	0.92
F3	.35**	.27**	1		3.55	0.83
F4	.51**	.43**	.33**	1	3.62	0.91

Note. Intercorrelations, M and SD are based on total Sample 1, N = 717.
**p < .01.

Reliability and Further Validation of the DWD-S (Study C)

In the context of a second questionnaire survey (Sample 2), reliability analysis as well as further construct validity of the final scale were investigated in Study C. Sample 2 is a panel study in which randomly selected students completed the scale at two different measurement times in order to calculate the retest reliability. The second survey (T2) took place after six weeks. In order to guarantee anonymity and still correctly assign the questionnaire pairs of the students from T1 and T2, a code procedure was used. Of the 134 students involved, n = 34 pairs of questionnaires could be correctly assigned. The Pearson correlation between the measured values at T1 and T2 was analyzed at the level of the items, factors, and the overall scale. The results for retest reliability show adequate results both at factor level (r = .63-.85) and about the overall scale (r = .70). All correlations have a p-value < .001 and are statistically significant taking the Bonferroni-Holm correction into account.

For the split-half coefficient, the items were sorted by item difficulty and discriminatory power and then assigned to the two parts of the test. Guttman's split-half coefficient is .78 (total Sample 1) and .80 (Sample 2). The total scale has a good internal consistency with α = .804 in total Sample 1 and with α = .820 in Sample 2. The EFAs were used to make a descriptive classification of homogeneous items based on theoretically defined characteristic dimensions. The structure found this way was checked via CFA based on new data sets and was confirmed, which also indicates a good construct validity of the scale (Pospeschill, 2010). In addition to the analysis of the characteristic structure of the items, correlations with other scales can also be used for construct validity.

For the analysis of convergent validity, an intercultural competence scale was used in Sample 1 (Reinders et al., 2011). As expected, the DWD-S shows medium to high correlations with the related intercultural competence and its dimensions (see Table 7). The correlation coefficients indicate the scale measuring a separate ability.

TABLE 7
Correlation between DWD-S and WIKI-KJ (total Sample 1, N = 717)

	DWD-S	F1	F2	F3	F4	*M*	*SD*	α
1. WIKI-KJ_total	.50**	.40**	.26**	.28**	.36**	3.24	0.51	.88
2. WIKI-KJ_contact	.21**	.16**	.15**	.08	.06	2.61	0.76	.79
3. WIKI-KJ_adaptivity	.51**	.41**	.24**	.32**	.43**	3.50	0.52	.76
4. WIKI-KJ_openness	.53**	.44**	.25**	.31**	.43**	3.44	0.60	.84

Note. DWD-S = Dealing With Diversity in Schools; WIKI-KJ = a German scale to measure intercultural competence of children and adolescents. 1-4 = correlation coefficient r; WIKI-KJ_total = total score intercultural competence; WIKI-KJ_contact = contact frequency with other-cultural peers; WIKI-KJ_adaptivity = adaptivity of other-cultural contact behavior; WIKI-KJ_openness = openness to intercultural contact (Reinders et al., 2011).
** p < .01.

In Study C, Sample 2 was used to examine correlations between the DWD-S and its factors with selected GMF scales to capture group-focused enmity (Krause & Zick, 2013). Group-focused enmity includes derogatory and exclusionary attitudes toward people solely based on their membership in a social group. Sample 2 shows clear negative correlations with homophobia, sexism, the devaluation of disabled people, and xenophobia (see Table 8). The lower the respondents' diversity acceptance, the more likely they

will devalue foreigners or minorities. Also, respondents with high diversity acceptance show less in-group favoritism (subscale GMF-EV). All the correlations found are coherent in terms of content, conform to expectations, can be interpreted meaningfully, and are also evident at the factor level.

TABLE 8
Correlation between DWD-S and group-focused enmity

	DWD-S	F1	F2	F3	F4	*M*	*SD*	α
1. GMF-HO	–.36**	–.22**	–.04	.13*	–.42**	1.01	0.53	.72
2. GMF-FF	–.52**	–.33**	–.24**	.19**	–.51**	1.44	0.85	.75
3. GMF-SX	–.40**	–.21**	–.06	.14*	–.47**	1.04	0.59	.75
4. GMF-BE	–.31**	–.08	–.19**	.03	–.37**	1.19	0.69	.72
5. GMF-ISL	–.51**	–.37**	–.21**	.18**	–.48**	1.49	0.95	.71
6. GMF-EV	–.38**	–.24**	–.16*	.14*	–.36**	1.77	0.98	.66

Note. DWD-S = Dealing With Diversity in Schools. 1-6 = group-focused enmity (GMF) short scale (Krause & Zick, 2013); GMF-HO = homophobia; GMF-FF = xenophobia; GMF-SX = sexism; GMF-BE = devaluation of disabled people; GMF-ISL = islamophobia; GMF-EV = privileges of the established, based on Sample 2, $N = 134$. *$p < .05$; **$p < .01$.

With total Sample 1, correlations between the DWD-S and its factors with the AIDA (Assessment of Identity Development in Adolescence) scale were examined (see Table 9). The AIDA scale (Goth et al., 2012) is a diagnostical test for psychiatric disorders, but it can also be used as a research instrument for measuring adolescent identity development. AIDA captures identity development regarding a personality function. It allows a dimensional differentiation between healthy identity integration and severe identity diffusion associated with a personality disorder risk. The AIDA total score represents the extent of identity diffusion, whereby the two major scales, discontinuity, and incoherence, are each differentiated into three subscales, which are not discussed further here. Since high scores on the AIDA scale are associated with difficulties and deficiencies in individual identity development, negative correlations between the DWD-S and AIDA were expected. Accordingly, students with high scores in identity diffusion would react more negatively to foreignness or otherness. The expected negative correlations can be seen in Table 9: the higher the diversity acceptance, the lower the expressions of identity diffusion, with correlations also appearing at the level of the two main scales, discontinuity, and incoherence.

DISCUSSION

Complex individual abilities such as diversity acceptance (but also empathy and the readiness for change) are receiving increased attention nowadays. They correspond most closely to the demands and challenges of the modern world and are, therefore, increasingly important in the school context. Due to their hybrid characteristics, they blur the distinction between competencies and personality traits: on the one hand, they are skills that are necessary to solve concrete challenges, and on the other hand, they are complex and profound abilities that are expected to be difficult to change (Pietzonka, 2021a). Skills for dealing with

TABLE 9
Correlation between the DWD-S and the AIDA scales

	DWD-S	F1	F2	F3	F4	*M*	*SD*	α
1. Diffusion	–.18**	–.14**	–.12**	–.24**	–.07	52.81	11.37	.99
2. Discontinuity	–.17**	–.13**	–.16**	–.18**	–.08*	52.79	11.39	.96
3. Incoherence	–.17**	–.14**	–.08	–.26**	–.06	52.82	11.26	.98

Note. DWD-S = Dealing With Diversity in Schools; AIDA = Assessment of Identity Development in Adolescence. 1-3 = AIDA scales (Goth et al., 2012); T-values were considered for all evaluations. 1 = AIDA total score (identity diffusion), 2 = AIDA/main scale discontinuity, 3 = main scale incoherence, based on total Sample 1, $N = 717$.
*$p < .05$; **$p < .01$.

diversity are individually essential resources for achieving one's goals in different social contexts and situations in an appreciative manner. Diversity acceptance, in particular, could play a key role here: the measurement of this ability captures the handling of diversity across dimensions and thus represents a broader ability than measurement procedures that only refer to individual diversity dimensions (Pietzonka, 2021b). In the context of this paper, the DWD-S scale adaptation was developed as a first approach to making this ability measurable among adolescents in the school context. The adapted scale has different potentials in the context of the diverse challenges of educational-psychological contexts. In adapting the scale for the school context, an attempt was made to align the items as closely as possible to the adult scale. The group of authors deliberately decided against creating new items. Instead, they used the item pool of the previous version of the adult scale, which was designed based on extensive literature research with the support of the expertise of experts and item analyses. Nonetheless, the items were modified and tested to suit young people, and the dimensionalization was also examined within the factor analyses. This effort was necessary because it could not be ruled out that other items from the item pool would "prevail" in the survey of the young people or that an alternative dimensionalization about the original scale would become apparent. However, within the framework of the EFA and CFA, a model that corresponds to the adult scale in terms of dimensionalization and factor names has prevailed. At first glance, this is not surprising; after all, an angler can only catch the fish that were previously put into the lake. Nevertheless, this circumstance also speaks for the relevance of diversity acceptance, since in an independent sample, with modified items, with a different population (adolescents) as well as changed framework conditions (school context), an almost similar structure in the response behavior is shown as in the initial scale. The EFA produced a factor model that was confirmed by a separate sample in the CFA. The model consists of four factors that can be meaningfully interpreted based on the items. The quality of the selected model can be rated as good with regard to the relevant parameters (RMSEA, SRMR, CFI, TLI). The first testing phase shows that the scale's quality criteria are sufficient to be used as a survey instrument.

Limitations

At first, the high number of outliers in the present study may sound surprising. Nevertheless, this is common in surveys with young people. Considering that multivariate outlier analyses with adult respondents regularly yield outlier rates of more than 10% (Hampel et al., 2005; Huber & Ronchetti, 2009), the

anomalies found in this study are a high but realistic value. It should be taken into account that some respondents were quite young students and that the survey was relatively time-consuming (111 items). In order to rule this out, a follow-up investigation of the response patterns of the excluded questionnaires was carried out, which revealed that the majority of the respondents concerned did not even attempt to deal with the content of the questions but instead ticked nonsense or decorative answer patterns, which speaks against a substantive engagement with the items and legitimizes the exclusion. The high outlier rate could indicate low compliance associated with low data quality.

Drawing representative samples was impossible, so no standard values could be derived for the DWD-S scale. Representative samples were considered desirable for the factor analyses but methodologically not necessary, especially because no substantial bias due to school-type-specific factors could be determined. Some of the items measuring diversity acceptance have a normative character and address, for example, bias against certain social groups. It can be assumed that these items are highly related to ideological or political beliefs and are thus influenced by the effects of social desirability. Most questionnaires were completed in a classroom context, so the survey setting could not be kept entirely anonymous. Accordingly, respondents may have chosen rather "politically correct" answers. These answers do not necessarily correspond to the actual beliefs of the respondents and represent a methodological limitation of the data collection.

The total number of items per factor is quite low. According to the central limit theorem of statistics, the available mean values are approximately normally distributed and, thus, also interval-scaled (Davidson, 2001) — but this only applies if the number of summands used to calculate the mean value is sufficiently large. The factors of the DWD-S are captured by three to four items, which corresponds to a small number of summands from a statistical point of view (Carifio & Perla, 2007). As a result, there may not be sufficient approximate interval scaling, which could have led to biased and possibly overestimated correlations. Furthermore, it does not follow from a high correlation that a linear relationship exists. A nonlinear relationship can be misinterpreted as a linear one due to a high correlation (Anscombe, 1973).

The question of whether an overall value can be calculated in the case of a nonrepresentative sample can also be discussed controversially. All factors measure one essential aspect of the ability. It is not plausible in terms of content that all factors F1 to F4 must have a high or low value to assign a high or low degree of diversity acceptance to a person. Instead, it can be assumed that, taken as a whole, all characteristics reflect the degree of diversity acceptance that exists. This implies the calculation of a total value. However, the contribution of each factor to the total is different. Thus, it makes sense to calculate the total value by means of the factors F1 to F4 weighted with the RVP values so that the relative variance explanations of the factors in calculating the total value are taken into account (Pietzonka & Kolb, 2021).

Practical Implications and Research Desiderata

In addition to researching young people's diversity acceptance, the scale shows potential for practical use. With a standardized scale, individual strengths and weaknesses could be recorded through an inventory, for example, in the context of school diversity training. As an instrument for schools, however, it is still rather inconvenient due to the inversely formulated items and the mathematical formula. An online-based and user-friendly version in which young people can fill out the scale directly is conceivable and would make the scale available to a wider public as an authentic self-assessment tool. The test phase shows that the quality criteria of the scale are sufficient so that it can be used as a survey instrument in research. The criterion validity is a research desideratum. It is therefore necessary to compare the scale with other scales, which also record the

individual dealing with social diversity in relation to a dependent criterion variable (e.g., discrimination intentions). This would be desirable to ensure the practical relevance of the DWD-S. The identification of the predictors of diversity acceptance is an important research desideratum. Questions about the development, influencing factors, and changeability of diversity acceptance can thus be investigated, which suggests valuable practical implications for diversity education in schools. This raises the question of how "stable" this ability is over time and situations. Whether and to what extent diversity acceptance can be changed with training is a research desideratum that can only be answered empirically. However, based on the above definition, it can be expected to be a stable trait-like ability that is relatively persistent over time. Schools are investing increasingly in changing skills for dealing with diversity, for instance, in the context of school projects, diversity training, or diversity workshops for students or teachers. In addition to dispositive factors, further research could examine socialization and educational processes in and out of school and individual maturation processes. For example, the type, characteristics, and duration of stays abroad and their possible influence on diversity acceptance could be examined with the new scale in a panel study. At this point, identity development could play a key role. Adolescence is a phase in which enormous cognitive and emotional changes occur. In addition to school and family, other social contexts (group affiliations, free time, social media, peers) play an important role that can influence the individual's approach to diversity. In addition to the published organization-related scales, the authors plan to develop a new scale that captures diversity acceptance without an organizational context. In addition, the scale was used in the meantime in a representative sample to derive norm values.

References

Anscombe, J. F. (1973). Graphs in statistical analysis. *American Statistician*, *27*(1), 17-21. https://doi.org/10.230/2682899

Bredella, L., Meißner, F. J., Nünning, A., & Rösler, D. (2000). Grundzüge einer Theorie und Didaktik des Fremdverstehens [Basic features of a theory and didactics of foreign understanding]. In L. Bredella, F. J. Meißner, A. Nünning, & D. Rösler (Eds.), *Wie ist Fremdverstehen lehr- und lernbar?* [How can understanding others be taught and learnt?] (pp. IX-LII). Narr.

Bredendiek, M. (2015). *Menschliche Diversität und Fremdverstehen — Eine psychologische Untersuchung der menschlichen Fremdreflexion* [Human diversity and understanding others — A psychological investigation of human reflection on others]. Springer. https://doi.org/10.1007/978-3-658-10313-2_1

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106-116. https://doi.org/10.3844/jssp.2007.106.116

Clark, R., Coleman, A. P., & Novak, J. D. (2004). Brief report: Initial psychometric properties of the everyday discrimination scale in black adolescents. *Journal of Adolescence*, *27*(3), 363-368. https://doi.org/10.1016/j.adolescence.2003.09.004

Clayton, J. K. (2010). Changing diversity in U.S. schools. *Education and Urban Society*, *43*, 671-695. https://doi.org/10.1177/0013124510380909

Davidson, J. (2001). *Econometric theory.* Blackwell Publishing. https://doi.org/10.1017/S0266466603004092

Diersch, N., & Walther, E. (2010). Umfrageforschung mit Kindern und Jugendlichen [Survey research with children and young people]. In E. Walter, F. Preckel, & S. Mecklenbräuke (Eds.), *Befragungen von Kindern und Jugendlichen: Grundlagen, Methoden und Anwendungsfelder* [Surveys of children and adolescents: Basics, methods and fields of application] (pp. 297-318). Hogrefe.

Ensign, K. A., Yiamouyiannis, A., White, K. M., & Ridpath, B. D. (2011). Athletic trainers' attitudes toward lesbian, gay, and bisexual National Collegiate Athletic Association student-athletes. *Journal of Athletic Training*, *46*(1), 69-75. https://doi.org/10.4085/1062-6050-46.1.69

Erikson, E. H. (1968). *Identity: Youth and crisis.* Norton & Co. https://doi.org/10.1002/bs.3830140209

Eurostat. (2022). *Migration and migrant population statistics.* Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migration_and_migrant_population_statistics

Filzmoser, P., & Hron, K. (2008). Outlier detection for compositional data using robust methods. *Mathematical Geosciences*, *40*, 233-248. https://doi.org/10.1007/s11004-007-9141-5

Goth, K., Foelsch, P., Schlüter-Müller, S., Birkhölzer, M., Jung, E., Pick, O., & Schmeck, K. (2012). Assessment of identity development and identity diffusion in adolescence — Theoretical basis and psychometric properties of the self-report questionnaire AIDA. *Child and Adolescent Psychiatry and Mental Health*, *6*(27). https://doi.org/10.1186/1753-2000-6-27

Graham, S. (2018). Race/ethnicity and social adjustment of adolescents: How (not if) school diversity matters. *Educational Psychologist*, *53*(2), 64-77. https://doi.org/10.1080/00461520.2018.1428805

Hampel, F. R., Ronchetyti, E. M., Rousseeu, P. J., & Stahel, W. A. (2005). *Robust statistics: The approach based on influence functions.* John Wiley & Sons. https://doi.org/10.1002/9781118186435

Hoffman, E., & Verdooren, A. (2019). *Diversity competence — Cultures don't meet, people do.* CAB International. https://doi.org/10.1079/9781789242409.0000

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424-453. https://doi.org/10.1037/1082-989X.3.4.424

Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics.* John Wiley & Sons. https://doi.org/10.1002/9780470434697

Hurrelmann, K., & Bauer, U. (2015). *Einführung in die Sozialisationstheorie* [Introduction to socialization theory]. Beltz.

Kiegelmann, M. (2009). Sprachenentwicklungspsychologische Voraussetzungen von Kindern und Jugendlichen und deren Konsequenz für die Kompetenz von Befragten [Language developmental psychological preconditions of children and adolescents and their consequence for the competence of interviewees]. In E. Walter, F. Preckel, & S. Mecklenbräuke (Eds.), *Befragungen von Kindern und Jugendlichen: Grundlagen, Methoden und Anwendungsfelder* [Surveys of children and adolescents: Basics, methods and fields of application] (pp. 33-44). Hogrefe.

Krause, D., & Zick A. (2013). Kurzskalen zur Gruppenbezogenen Menschenfeindlichkeit (GMF) [Short scales on group-focused enmity]. In C. J. Kemper & M. Zenger (Eds.), *Psychologische und sozialwissenschaftliche Kurzskalen: standardisierte Erhebungsinstrumente für Wissenschaft und Praxis* [Psychological and social science short scales: standardized survey instruments for science and practice] (pp. 119-121). Medizinisch Wissenschaftliche Verlagsgesellschaft. https://doi.org/ 10.1055/s-0034-1396798

Moosbrugger, H., & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion* [Test theory and questionnaire construction]. Springer. https://doi.org/10.1007/978-3-540-71635-8

Pena, D., & Prieto, F. J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, *43*, 286-310. https://doi.org/10.1198/004017001316975899

Pietzonka, M. (2016). Diversity-Kompetenz als Lernziel der Hochschulbildung? [Diversity competence as a learning goal of higher education?]. *HQSL*, *58*, 1-26.

Pietzonka, M. (2018). Schlüsselkompetenzen zum Umgang mit sozialer Vielfalt für die Arbeitswelt 4.0: Einordnung, Kennzeichnung und Messung [Key competences for dealing with social diversity for the working world 4.0: Classification, labelling and measurement]. In B. Hermeier, T. Heupel & S. Fichtner-Rosada (Eds.), *Arbeitswelten der Zukunft* [Working worlds of the future] (pp. 477-796). Springer-Gabler. https://doi.org/10.1007/978-3-658-23397-6_25

Pietzonka, M. (2021a). Fähigkeit zum Umgang mit Diversität und ihre Vermittlung in Hochschulen [Ability to deal with diversity and its mediation in higher education institutions]. In M. Gruber, K. Ogris, R. Wieser, & B. Brese (Eds.), *Diversität im Kontext Hochschullehre: Best Practice* [Diversity in the context of university teaching: best practice] (pp. 21-26). Waxmann.

Pietzonka, M. (2021b). Die Ratingskala DWD-O zur Messung von Diversitätsakzeptanz im organisatorischen Kontext [The DWD-O rating scale for measuring diversity acceptance in an organisational context]. *Zeitschrift für Arbeits- und Organisationspsychologie*, *65*(1), 17-26. https://doi.org/10.1026/0932-4089/a000340

Pietzonka, M., & Kolb, C. J. (2021). Diversity acceptance as an individual ability: The new rating scale DWD-O5 for the organizational context. *TPM — Testing, Psychometrics, Methodology in Applied Psychology*, *28*(4), 1-16. https://doi.org/10.4473/TPM28.4.1

Pike, C. K. (2002). Measuring racial climate in schools of social work: Instrument development and validation. *Research on Social Work Practice*, *12*(1), 29-46. https://doi.org/10.1177/104973150201200104

Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation* [Test theory, test construction, test evaluation]. UTB. https://doi.org/10.1515/infodaf-2012-2-375

Reinders, H., Gniewosz, B., Gresser, A., & Schnurr, S. (2011). Erfassung interkultureller Kompetenzen bei Kindern und Jugendlichen. Das Würzburger Interkulturelle Kompetenz-Inventar (WIKI-KJ) [Assessment of intercultural competences in children and adolescents. The Würzburg Intercultural Competence Inventory]. *Diskurs Kindheits- und Jugendforschung*, *4*, 429-452.

Streitmatter, J. L., & Pate, G. S. (1989). Identity status development and cognitive prejudice in early adolescents. *The Journal of Early Adolescence*, *9*, 142-152.

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup relations. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-48). Brooks/Cole.

Terhart, E. (2015). Heterogenität der Schüler — Professionalität der Lehrer: Ansprüche und Wirklichkeiten [Heterogeneity of students — Professionalism of teachers: Demands and realities]. In C. Fischer, M. Veber, C. Fischer-Onstrup, & R. Buschmann (Eds.), *Umgang mit Vielfalt – Aufgaben und Herausforderungen für die Lehrerinnen- und Lehrerbildung* [Dealing with diversity - tasks and challenges for teacher training] (pp. 13-30). Waxmann.

Timmermanns, S. (2007). Dealing with diversity in education and counselling: Lesbians and gays at German schools. In L. van Dijk & B. van Driel (Eds.), *Challenging homophobia. Teaching about sexual diversity* (pp. 47-59). Trentham Books.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying contest effects in attidude measurement. *Psychological Bulletin*, *103*, 229-314.

Walter, C., & Schmidt, B. (2016). Paradigmenwechsel in der Gesellschaft: Wie schützt Diversity Kompetenz vor Diskriminierung? [Paradigm shift in society: How does diversity competence protect against discrimination?]. In P. Genkova & T. Ringeisen (Eds.), *Handbuch Diversity Kompetenz* [Diversity competence handbook] (pp. 235-249). Springer. https://doi.org/10.1007/978-3-658-08594-0_20