Praetorius, Anna-Katharina; Drexler, Katharina; Rösch, Lisa; Christophel, Eva; Heyne, Nora; Scheunpflug, Annette; Zeinz, Horst; Dresel, Markus

Judging students' self-concepts within 30 seconds? Investigating judgement accuracy in a zero-acquaintance situation

Judging Students' Self-Concepts Within 30 Seconds? Investigating Judgment Accuracy In A Zero-Acquaintance Situation

Anna-Katharina Praetorius & Katharina Drexler

University of Augsburg, Germany

Lisa Rösch

University of Heidelberg, Germany

Eva Christophel

University of Koblenz-Landau, Germany

Nora Heyne

University of Koblenz-Landau, Germany

Annette Scheunpflug

University of Bamberg, Germany

Horst Zeinz

University of Muenster, Germany

Markus Dresel

University of Augsburg, Germany

Author Note

Anna-Katharina Praetorius, Katharina Drexler, and Markus Dresel, Department of Psychology, University of Augsburg, Germany

Lisa Rösch, Department of Psychology, University of Heidelberg, Germany

Eva Christophel, Department of Educational Psychology, University of Koblenz-Landau, Germany

Nora Heyne, Department of Teacher Education, University of Koblenz-Landau, Germany

Annette Scheunpflug, Department of Education, University of Bamberg, Germany

Horst Zeinz, Department of Education, University of Münster, Germany

Abstract

Judging students' academic self-concepts accurately is assumed to be a necessary condition for adaptive classroom instruction. Previous studies found moderate correlations between teachers' judgements and the self-reported self-concepts. Reference values for interpreting these correlations are missing. In the present study, self-other agreements with a zero-acquaintance premise (see Ambady & Rosenthal, 1992) were proposed as such a reference value.

A total of 88 teachers judged the self-concepts of their own students. Additionally, 97 persons judged students' self-concepts based on 30-second videos of the students. Intra-individual correlations were calculated between judgements and actual self-concepts.

The analyses showed that agreements in the natural classroom setting and in a zero-acquaintance setting both displayed generally moderate correlations and did not differ significantly from one another.

Based on the results, the accuracy of teachers in judging students' self-concepts is considered to be low. Possible reasons for the low accuracy of teachers' judgement accuracy are discussed.

1.   Introduction

Diagnosing student characteristics is an important task, a task teachers are expected to execute accurately. Previous research has shown that teacher-student agreements regarding students' motivational characteristics usually prove to show mid-range correlations. However, to date it is unclear whether these agreements should be interpreted as indicators of low, medium, or high accuracy. In the present study, self-other agreements in a zero-acquaintance premise are proposed as an aid to interpret and classify the correlations found between teacher judgements and the respective student characteristics.

1.1   Teacher judgements and their accuracy

According to current views of teaching (e.g., constructivist views of teaching for understanding, see Cohen, 1993; Pauli, Reusser, & Grob, 2007), high quality instruction has to be tailored to the specific students being taught. A teacher might, for example, be very good at describing complex issues from an objective point of view. However, if the students do not have the requisite basic knowledge in the issue at hand, they will not be able to follow. A teacher might also present his or her students with challenging tasks which they would generally be able to solve with a good deal effort. However, if the students are convinced that they are not able to solve the tasks (e.g., due to low self-concepts), they will not work diligently enough to complete them. In both cases, the teacher needs to adapt his or her

classroom activities to the learning prerequisites of the students (e.g., Begeny, Eckert, Montarello, & Storie, 2008; Edelenbos & Kubanek-German, 2004; Hoge & Coladarci, 1989). In order to be able to do so, teachers need skills which allow them to accurately judge their students' learning prerequisites.

Previous research has shown that the ability of teachers to carry out accurate judgements varies considerably for different student characteristics: According to a meta-analysis conducted by Südkamp, Kaiser, and Möller (2012), school achievement is judged with a mean correlation of $r = .63$ between school achievement measured with standardized achievement tests and the respective teacher judgements. Other student characteristics are judged considerably less accurately by teachers: For example, analyses by Boehnke, Silbereisen, Reynolds, and Richmond (1986) revealed that correlations of teacher judgements of students' achievement anxiety and the self-reported achievement anxiety were rather poor on average ($r = .21$). Spinath (2005) analysed the judgement accuracy of teachers regarding the academic self-concepts of students, as well as their learning motivation, and found median correlations of $r = .39$ and $r = .20$.

1.2 Interpreting the amount of judgement accuracy

Summarizing the existing studies on teachers' judgement accuracy, teachers are considerably less accurate in judging motivational student characteristics when compared to judgements of student achievement. These deviations can be explained by differences in the detection and utilization of information relevant for inferences of the respective student characteristic (see Funder, 1995). There are many opportunities to assess student achievement in a classroom context (e.g., based on the correctness of the written and oral statements made by students). At the same time, there are many situations in which students may openly express their motivational status. However, since the primary focus of school is on achievement, it is conceivable that teachers pay more attention to achievement-related information than to motivation-related information, and thus detect more achievement-related information. Even if teachers do pay attention to motivation-related information, it is likely that they do not only rely on valid information for their judgements, as the actions and the achievement of students do not necessarily correspond to their motivation, for example their academic self-concepts (Givvin, Stipek, Salmon, & MacGyvers, 2001; Praetorius, Berner, Zeinz, Scheunpflug, & Dresel, 2013).

If one agrees that judging motivational characteristics is a difficult task, how can the existing correlations between teacher judgements and student motivational characteristics be

interpreted with regard to high or low judgement accuracy? Presently there is a lack of clear interpretation guidelines. Some authors regard the existing correlations to indicate considerable precision (e.g., Marsh & Craven, 1991), while other authors consider them as an indicator of low judgement accuracy (e.g., Spinath, 2005). One possibility to classify the amount of teacher-student agreements is presented by the conventions proposed by Cohen (1992). According to these conventions, correlations of around $r = .10$ are considered to be small, correlations close to $r = .30$ are considered to be moderate and those at or exceeding $r = .50$ are considered to be large. However, these are general rules of thumb and they have to be adjusted to each specific characteristic under consideration. A specific reference value to interpret existing teacher-student correlations would thus be very useful. One approach to empirically identify such a reference value is the zero-acquaintance approach (see also Spinath, 2005).

1.3 The zero-acquaintance approach and its application to judgements of students' motivational characteristics

A common belief is that accurate judgements of persons require large amounts of both interpersonal interaction and information about the person to be judged. With the zero-acquaintance or thin-slices of behaviour paradigm (for an overview see Ambady, Bernieri, & Richeson, 2000) it could, however, be shown that persons can judge strangers surprisingly well after only short observations (usually less than 5 min), and without personal interaction. In many studies (see e.g., Ambady & Grey, 2002; Carney, Colvin, & Hall, 2007; Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004), the zero-acquaintance situation is established by presenting persons with short video clips of unacquainted targets. Judges are then asked to rate the targets with regard to certain characteristics, for example personality traits like extraversion.

In Social, Personality as well as Clinical Psychology, many studies have demonstrated that judgements based on such very brief observations lead to judgement accuracies at levels significantly above chance (Ambady, Krabbenhoft, & Hogan, 2006; Ambady & Rosenthal, 1992). In their meta-analysis, Ambady and Rosenthal (1992) summarized the results of 38 studies focusing on the predictive validity of judgements based on minimal information. The studies covered several judgement topics, for example anxiety or depression levels as well as voting behaviour. On average, the correlation between a judgement and an objective criterion reached $r = .39$ ($r_{min.} = .11$, $r_{max.} = .87$). Neither the length of the observation period (20 s up to 5 min), nor the kind of behavioural information (e.g., face or body) nor the setting (laboratory vs. natural) moderated the effects. The existing correlations are mainly explained

through the non-verbal cues (e.g., facial expression) the judged persons exhibit (Ambady et al., 2000).

Some zero-acquaintance studies have also investigated persons in the school context (e.g., Ambady & Rosenthal, 1993; Babad, Avni-Babad, & Rosenthal, 2003; Pretsch, Flunger, Heckmann, & Schmitt, 2013). However, these studies focused exclusively on teachers' instructional practices or their personality (e.g., to what extent a few seconds of teachers' non-verbal behaviour can predict students' ratings of teachers in Babad et al., 2003). In the present study, the video-based zero-acquaintance approach is transferred to teacher judgements of a motivational student characteristic, namely the students' academic self-concepts.

The academic self-concept is defined as the student's self-perception of his or her academic abilities. Pivotal, regarding self-concepts, is that these perceptions of the self are subjective. They thus can be realistic (i.e., in agreement with the actual abilities), but they do not have to be. The academic self-concept is considered to be one of the most important motivational learning prerequisites of students (see Möller, Pohlmann, Köller, & Marsh, 2009). This is due to the fact that, among others, academic self-concepts have a considerable influence on academic behaviour and, ultimately, achievement (see for example the meta-analysis of Huang, 2011).

Teachers' judgement accuracies regarding the academic self-concepts of students usually range – according to the conventions of Cohen (1992) – between small and moderate correlations (e.g., Givvin et al., 2001; Marsh, Smith, Barnes, & Butler, 1983; Spinath, 2005), even though there is a large variation between studies ($r_{min.}$ = -.10, Zirkel & Gable, 1977; $r_{max.}$ = .67, Chang, 1976).

1.4 Research question and hypothesis

To obtain a reference value for interpreting the agreement between teacher judgements of students' self-concepts and students' self-reported self-concepts in classroom settings (i.e., correlations found in previous research), the present research compares the accuracy of teacher judgements in such an instructional setting with the accuracy of judgements based on minimal information about the targets.

Several studies have shown that the agreement between judges and the self-view of the persons being judged is higher with longer acquaintance (e.g., Biesanz, West, & Millevoi, 2007; Funder & Dobroth, 1987; Watson, Hubbard, & Wiese, 2000). Teachers interact with their students on a daily basis, thus they should know them well. We therefore hypothesized that the average judgement accuracy of teachers assessing the academic self-concepts of their

students in natural classroom settings should be considerably higher than the judgement accuracy of persons based on minimal information.

2.   Method

2.1   Participants

The data used for analyses of teachers in a natural classroom setting was collected in the 2009/10 school year (see Praetorius et al., 2013). The schools investigated were 19 public intermediate secondary schools ("Realschulen")[1] in the German state of Bavaria. The current analysis was conducted for Mathematics and German teachers for whom full data sets were available. Altogether, data were analysed for a total of 88 teachers (45% female; $M = 9.15$ years of professional experience, $SD = 9.25$; min. = 1; max. = 35).

Four groups of judges were tested in the zero-acquaintance condition in order to gain information about the generalizability of the results: Group One consisted of 30 undergraduate psychology students (26 female; $M = 21.32$ years, $SD = 2.91$) attending a German university. They participated in exchange for course credit. Group Two were 35 teaching students (24 female; $M = 23.63$ years, $SD = 7.30$) who also participated in exchange for course credit. Group Three consisted of 18 school counselling students (five female, 11 without information; $M = 24.00$ years, $SD = 1.63$). These students participated in the context of a course lecture. Group Four were 16 teachers employed at primary and secondary schools (13 female; $M = 34.50$ years, SD = 9.71). They received a 10 Euro voucher for their participation. The sampling in the zero-acquaintance condition was incidental for all investigated groups.

2.2   Stimuli

For the natural classroom sample, 1280 students (48% female; $M = 13.74$ years; $SD = 1.64$) in grades five through nine served as targets. The students were taught by those teachers under investigation for judgement accuracy. Entire classroom samples were investigated. Participation was limited to those students for whom parental permission was given. The number of students participating per class ranged between seven and 32 ($M = 20.53$).

In the zero-acquaintance condition students working on a text-picture task served as targets (see Baadte, Christophel, Heyne, & Schnotz, 2013). The text and picture comprehension tasks required that they answer six multiple choice questions (Schnotz et al., 2010). The students were asked to think aloud while solving the tasks. Of the 85 participating

---

[1] There are three secondary school tracks in the German school system: a lower, an intermediate, and an academic track.

students, eight (five female), in grades five to seven representing all secondary school tracks in the German school system, were chosen to serve as targets in the present investigation. The students were selectively chosen from the entire catalogue of video material (see Heyne, Oswald, Baadte, & Schnotz, 2010). The criterion for selection was meaningful video material (e.g., some of the students did not succeed in thinking aloud). Every student was filmed from a frontal perspective while working on a different task. After solving the tasks, the students completed a self-concept questionnaire. In total, the film sequences were 6 to 12 min in length (Christophel, 2014). The sequences were edited down to 30 s as this duration is common in zero-acquaintance studies (e.g., Ambady & Gray, 2002; Ambady & Rosenthal, 1993; Back, Schmukle, & Egloff, 2010). To obtain comparable video sequences, the moment that the student started to read the task aloud was set as the starting point of the video.

2.3 Procedure

For the natural classroom sample, student data was collected in the classroom. The teacher questionnaires were distributed at the same time as student data were collected and then returned individually by the teachers.

In the zero-acquaintance condition, each judge was provided with a computer and headphones and independently judged the eight students. The order of the video sequences was randomized. In each case, the judge watched the video sequence and judged the academic self-concept of the student back-to-back. The 30-second videotapes of the students thinking aloud while working on a task was the only information about the students available to the judges.

2.4 Instruments

2.4.1 Academic self-concepts of the students

For the natural classroom sample, the domain-specific self-concepts of the students were assessed using two scales taken from the PISA 2000 survey (see Kunter et al., 2003). They are formulated for the subjects of German ($\alpha = .83$; sample item: "I am a hopeless case in the subject of German") and mathematics ($\alpha = .89$; sample item: "I am always good at maths"). Each of the two scales comprised three items and used a Likert-type scale ranging from 1 (*not true*) to 4 (*true*).

The academic self-concepts of the students for the zero-acquaintance condition were assessed with 5 items from the SESSKO questionnaire (Schöne, Dickhäuser, Spinath, & Stiensmeier-Pelster, 2002). A sample item is: "I am not talented/ very talented for school". The scale was bipolar and ranged from 1 to 5. The internal consistency was $\alpha = .95$.

2.4.2 Judgements of students' self-concepts.

In the natural classroom sample, teachers were asked to judge the domain-specific self-concepts of their students in the subject they teach. The following question was used to assess the teacher judgements: "How does the student evaluate his/ her abilities in mathematics/ German?". Judgements were assessed with a four point Likert-type scale (1 = low; 4 = high). As a basis for their judgements, teachers were shown the students' self-concept items.

In the zero-acquaintance samples, judges were asked to evaluate each item as to how they think the student had answered the respective item. Analogous to the self-concept questionnaire, the answer scale used five points. The internal consistencies were satisfactory for all samples (α = .91 for the psychology students; α = .94 for the teaching students; α = .89 for the counselling teacher students; α = .96 for the teachers).

2.5   Analyses

Based on the criticism of Cronbach (1955) regarding the calculation of judgement accuracy, Helmke and Schrader (1987) proposed three components for assessing the judgement accuracy of teachers: A rank-order component, a level component, and a component of differentiation. The rank-order component is assumed to be the most important of the three.

In calculating the rank-order component, first, the means of the scale for every student and every judgement of a student were calculated. In the classroom sample, teachers' global judgements for the individual students were used for analyses. Second, for each judge, Pearson correlations were calculated between the judgements they made and the self-reported academic self-concepts of the students, i.e. intra-individual correlations were calculated for each of the judges over the targets judged. As the scale of measurement for regular correlation coefficients is not metric, these coefficients cannot be averaged, thus they were subsequently transformed to Fisher's *Z*, averaged across judges, and then transformed back to Pearson correlations.

In calculating the level component, the mean of students' self-concepts was subtracted from the mean of the judgements of a judge. The value zero thus indicates an optimal judgement as mean judgement and mean target value are equivalent. Values greater than zero indicate an overestimation of the level, values less than zero indicate an underestimation of the level.

In calculating the component of differentiation, the variation of the judgements of a judge was divided by the variation of the students' self-concepts. The value one thus indicates

an optimal judgement, as the variation of the judgements of a judge and the variation of the target values are identical. Values greater than one indicate an overestimation of the differentiation, values less than one indicate an underestimation of the differentiation.

3. Results

Table 1 presents the mean scores and standard deviations for the students' self-concepts and the judgements of these self-concepts.

The rank-order agreement between the academic self-concepts of the students and the respective judgements for the teachers in the classroom setting as well as all investigated

Table 1
Descriptive statistics for students' self-concepts and judgements of self-concepts.

	M	SD
Natural classroom sample		
Academic self-concept of students (n = 1280)	2.67	0.37
Judgements of their n = 88 teachers	2.52	0.29
Zero-acquaintance samples		
Academic self-concept of students (n = 8)	3.62	1.14
Judgements of n = 30 psychology students (Sample 1)	3.64	0.36
Judgements of n = 35 teaching students (Sample 2)	3.45	0.86
Judgements of n = 18 counselling teacher students (Sample 3)	3.46	0.88
Judgments of n = 16 teachers (Sample 4)	3.46	0.37

groups of judges in the zero-acquaintance condition can be found in Table 2. For all samples, the mean correlations were of moderate size ($.29 \leq r \leq .39$). In order to compare the correlations found in the classroom condition and those found in the zero-acquaintance conditions, the zero-acquaintance samples were merged into one sample. On a descriptive level, the mean zero-acquaintance correlation (r = .35) was even higher than the one found for the classroom setting (r = .29). Applying a test for two correlations in independent samples (see Steiger, 1980), the two correlations did, however, not differ significantly from each other (z = .60; $z_{5\%}$ = 1.65). Our hypothesis that the judgement accuracy of teachers in natural class settings is higher than judgement accuracy of persons in zero-acquaintance situations must, therefore, be rejected with regard to rank-order accuracy.

Calculating the level component revealed that teachers and university students, in all investigated samples, overestimated the mean level of the students' self-concepts on a descriptive level (see Table 2). Testing these descriptive differences against the optimal value zero using one-sample t-tests with a testing value of zero and a one-sided p-value revealed

that the teachers in the natural classroom sample significantly overestimated the mean level of their students' self-concepts, as did two of the four zero-acquaintance samples.

The analyses regarding the component of differentiation showed that, with one exception, all samples overestimated the variation of the students' self-concepts on a descriptive level (see Table 2). Testing these descriptive differences with regard to the variance homogeneity for variances in correlated samples (see Pitman, 1939) of the judgements and the self-concepts revealed that the overestimation was significant for the natural teacher sample as well as for two of the four zero-acquaintance samples.

4.   Discussion

When we meet someone for the first time, we immediately form an initial impression of this person. The longer we know him or her, the better we should know the person. Several studies have shown that, indeed, the agreement between judges and the self-view of the judged persons is higher with longer acquaintance (e.g., Biesanz et al., 2007; Funder & Dobroth, 1987; Watson et al., 2000). Teachers interact with their students on a daily basis, thus they should know them well. The results of the present study indicate, however, that this is not the case – at least regarding students' academic self-concepts. Teachers judge the academic self-concepts of their students with regard to all three accuracy components no more accurately than persons who have seen the students for a mere 30 s. Based on the knowledge that self-other agreements in the size of moderate correlations can possibly be achieved within such a short period of time, the measure of teachers' judgement accuracy regarding the academic self-concepts of their own students can only be interpreted as low.

According to Marsh and Craven (1991), teachers should be able to judge their students' self-concepts accurately as they have (a) a broad data basis regarding their students due to frequent interactions and (b) a broad frame of reference due to their knowledge of many different students. The authors concluded, based on the mainly moderate correlations between teacher judgements and students' self-concepts they found in their study, that the amount of self-other agreement is reasonably high. Based on the results of the research at hand we, in contrast, would conclude that the teacher-student agreement found for teachers is not satisfactory, as a similar extent of agreement was found for judgements based on 30-second observations. The mere existence of large data bases and broad frames of reference obviously do not necessarily mean that this information is used by teachers.

There are at least two possible explanations for this result. Both explanations are based on the assumption that students' motivational characteristics are not perceived as highly relevant by teachers: Many different things require the daily attention of teachers at school,

where the main objectives are insuring and assessing scholastic achievement. Thus, teachers probably consider motivational characteristics to be less important than academic achievement. Time constraints on teachers might restrict their capacity to focus attention on the identification of motivational characteristics (explanation 1). According to models of social information processing (e.g., Chaiken & Trope, 1999), a first step in information processing requires perception and focussing attention. It can be assumed that, primarily, the information which is perceived and focused on is that which is available and perceived as highly relevant. In a next step, only this kind of information should then be judged accurately. Experimental studies in the field of teachers' judgement accuracy support these assumptions (see Dünnebier, Gräsel, & Krolak-Schwerdt, 2009; Glock, Kneer, & Krolak-Schwerdt, 2011). Also, it may well be that the perceived subordinate importance of students' self-concepts results in heuristic decision-making (Tversky & Kahneman, 1973) when teachers are asked to judge their students' self-concepts (explanation 2). Such heuristics are time-saving and correct to at least some degree. One possible heuristic is the availability heuristic which implies that people mainly use information which is easily available when making a decision. In the case of judging students' self-concepts, easily available information comes in the form of their achievements. The differentiation between self-concept and achievement is a complex task as the two constructs are indeed interrelated to a considerable degree (e.g., mean $r = .57$ in a study of Marsh, 1992). First evidence regarding the hypothesis that teachers chiefly rely on the achievement of students when judging their academic self-concepts was found in a study conducted by Praetorius, Greb, Lipowsky, and Gollwitzer (2010).

One implication of the considerations mentioned above is that the similarities between the judgement accuracy of teacher judgements and the 30-second judgements should also be found for other motivational characteristics in that the arguments which arose apply to these characteristics as well. Regarding student achievement, however, it can be assumed that all teachers perceive this student characteristic to be more important than others and thus focus their attention on it to a larger degree. Teachers' judgement accuracy should therefore be considerably higher than the judgement accuracy found in zero-acquaintance studies. However, further empirical studies are necessary to clarify whether this assumption is true. Regardless of whether it is true or not the fact that, within 30 s, students' academic self-concepts can be judged with $.31 \leq r \leq .39$ questions whether correlations between teacher judgements and achievement of $r = .63$, as found in the meta-analysis of Südkamp et al. (2012), reflect high judgement accuracy.

Table 2
Accuracy components for all samples

Group	Intra-individual correlations				Level component			Component of differentiation		
	M	*SD*	*Min*	*Max*	*M*	*SD*	*t (p)*	*M*	*SD*	*t (p)*
Natural classroom sample										
Judgements of n = 88 teachers	**.29**	.34	−.45	.81	**0.13**	0.36	3.39 (.001)	**1.08**	0.37	2.38 (.01)
Zero-acquaintance samples										
Judgements of *n* = 30 psychology students (Sample 1)	**.39**	.40	−.51	.88	0.09	1.82	0.27 (.39)	**0.73**	0.14	8.19 (.000)
Judgements of *n* = 35 teaching students (Sample 2)	**.31**	.44	−.76	.91	**0.18**	0.36	2.96 (.003)	1.59	0.68	1.73 (.05)
Judgements of *n* = 18 counseling teacher students (Sample 3)	**.31**	.57	−.83	.84	**0.21**	0.29	3.07 (.003)	1.84	2.16	1.10 (.14)
Judgements of *n* = 16 teachers (Sample 4)	**.35**	.30	−.17	.70	0.16	0.37	1.73 (.05)	**1.55**	0.62	5.51 (.000)

Note. All correlations are Pearson correlations. The mean values were Fisher-Z transformed before averaging and transformed back afterwards.

The p-values for the t-tests are one-sided.

4.1 Limitations and further directions

The samples used in the present study were collected in one country, incidentally, and only considered university students and teachers. It thus remains unclear to what extent the findings of the present study can be generalized across the samples investigated.

One limitation regarding the interpretation of teachers in natural settings having low judgement accuracy is that different instruments were used in the natural and the zero-acquaintance settings. Students and raters were provided with similar instruments which used several items in the zero-acquaintance condition; whereas the teachers in the natural setting were only provided with a single item. The task for the teachers thus was more unspecific and the use of a single item, in all likelihood, raises unreliability. Both of these aspects may have lowered the magnitude of the teacher–student correlations in the natural setting. However, as the correlations in the natural setting were, on a descriptive level, even lower than those in the zero-acquaintance settings, it is highly unlikely that the higher unreliability and the lower specificity of the teacher ratings are the only reasons for the low teacher judgement accuracy found.

What remains unclear, based on the results of the present study, are the mediating processes that lead to the moderate correlations between judgements and self-concepts. On which cues do persons base their judgements if they observe students only for a very short period of time? One possibility to investigate this question is furnished by Brunswik (1956). This model allows one to examine which cues are used by persons in the judgements they make (cue utilization) and which cues are actually related to the characteristic being judged (cue validity). In assessing cue utilization and cue validity, observable cues are measured and subsequently correlated to the judgement as well as to the actual characteristic. A comparison between these correlations then allows for an estimation of the degree to which judges factor valid cues into their judgements.

The study at hand is a first attempt at transferring the zero-acquaintance approach to the field of teachers' judgement accuracy. Further studies should aim to strengthen the zero-acquaintance setting, for example by asking judges to rate the target trait as well as a non-target trait, or by varying the amount of time or the kind of behavioural information the targets come up against.

4.2   Conclusions

As a conclusion, the zero-acquaintance approach seems to be promising for research on teachers' judgement accuracy as the results of the present research can be used to interpret teachers' judgement accuracy regarding motivational student characteristics among students. The similarities among findings for all investigated samples point towards the generalizability of the results and underline the usefulness of transferring the zero-acquaintance approach to research on teachers' judgement accuracy. All in all, the results underscore the importance of enhancing teachers' judgement accuracy in assessing motivational characteristics such as students' self-concepts.

References

Ambady, N., Bernieri, F., & Richeson, J. (2000). Towards a histology of social behavior: Judgmental accuracy from thin slices of behavior. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, Vol. 32* (pp. 201–272). San Diego, CA: Academic Press.

Ambady, N., & Gray, H.-M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83*, 947–961. doi:10.1037//0022-3514.83.4.947

Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology, 16*, 4–13. doi:10.1207/s15327663jcp1601_2

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274. doi:10.1037/0033-2909.111.2.256

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431–441. doi:10.1037/0022-3514.64.3.431

Baadte, C., Christophel, E., Heyne, N., & Schnotz, W. (2013). Diagnostische und didaktische Kompetenz von Lehrkräften zur Förderung der Text-Bild-Integrationsfähigkeit bei Schülerinnen und Schülern der Sekundarstufe I – Schlussbericht zum Projekt DIKOL [Diagnostic and didactic skills of teachers in promoting text-image integration capacities among pupils in lower secondary education – Final report on the DIKOL project]. Landau, Germany: Universität Koblenz-Landau, Campus Landau.

Babad, E., Avni-Babad, D., & Rosenthal, R. (2003). Teachers' brief nonverbal behaviors in defined instructional situations can predict students' evaluations. J*ournal of Educational Psychology, 95*, 553–562. doi: 10.1037/0022-0663.95.3.553

Back, M. D., Schmukle, S. C., & Egloff, B. (2010). Why are narcissists so charming at first sight? Decoding the narcissism-popularity link at zero acquaintance. *Journal of Personality and Social Psychology, 98*, 132–145. doi:10.1037/a0016338

Begeny, J. C., Eckert, L. E., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly, 23*, 43–55. doi:10.1037/1045-3830.23.1.43

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self–other

agreement in judgments of personality. *Journal of Personality and Social Psychology, 92*, 119–135. doi:10.1037/0022-3514.92.1.119

Boehnke, K., Silbereisen, R. K., Reynolds, C. R., & Richmond, B. O., (1986). What I think and feel – German experience with the revised form of the children's manifest anxiety scale. *Personality and Individual Differences, 7*, 553–560. doi:10.1016/0191-8869(86)90132-7

Borkenau, P., Maurer, N., Riemann, R., Spinath, F.M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599–614.

Brunswick, E. (1956). *Perception and the representative design of experiments*. Berkeley, CA: University of California Press.

Carney, D.R., Colvin, C.R., & Hall, J.A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*, 1054–1072.

Chaiken, S. & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York, NY: Guildford Press.

Chang, T. S. (1976). Self-concepts, academic achievement, and teacher's rating. *Psychology in the Schools, 13*, 111–113.

Christophel, E (2014). *Lehrerfeedback im individualisierten Unterricht. Spannungsfeld zwischen Instruktion und Autonomie [Teacher feedback in individualized instruction. Striking the balance between instruction and autonomy]*. Wiesbaden, Germany: Springer.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. doi: 10.1037/0033-2909.112.1.155

Cohen, D. K. (1993). *Teaching for understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.

Cronbach, L. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin, 52*, 177–193. doi:10.1037/h0044919

Dünnebier, K., Gräsel, C., & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. Eine experimentelle Studie zu Ankereffekten [Judgment biases in scholastic performance assessment. An experimental study on anchor effects]. *Zeitschrift für Pädagogische Psychologie, 23*, 187–195. doi:10.1024/1010-0652.23.34.187

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of "diagnostic competence". *Language Testing, 21*, 259–283. doi:10.1191/0265532204lt284oa

Funder, D. C. (1995). On the accuracy of personality judgments: A realistic approach. *Psychological Review, 102*, 652–670. doi:10.1037/0033-295X.102.4.652

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52*, 409–418. doi:10.1037/0022-3514.52.2.409

Givvin, K. B., Stipek, D. J., Salmon, J. M., & MacGyvers, V. L. (2001). In the eyes of the beholder: Students' and teachers' judgments of students' motivation. *Teaching and Teacher Education, 17*, 321–331. doi: 10.1016/S0742-051X(00)00060-3

Glock, S., Kneer, J., & Krolak-Schwerdt, S. (2011). Impression formation or prediction? Category fit and task influence forensic person memory. *Journal of Forensic Psychology Practice, 11*, 391–405. doi:10.1080/15228932.2011.588529

Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education, 3*, 91–98.

Heyne, N., Oswald, E., Baadte C., & Schnotz W. (2010, August). Supporting integrative text and picture comprehension by feedback. *Poster presented at the EARLI SIG 2 Conference on Comprehension of Text and Graphics, Tübingen, Germany*. Abstract retrieved from http://www.earlisig2-2010.de/sig2-2010/program.php

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*, 197–313. doi:10.3102/00346543059003297

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*, 505–528. doi: 10.1016/j.jsp.2011.07.001

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., et al. (2003). *PISA 2000: Dokumentation der Erhebungsinstrumente [PISA 2000: Documentation of the survey instruments]* Vol. 72. Berlin, Germany: Max-Planck-Institut für Bildungsforschung.

Marsh, H. W. (1992). Content specifity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology, 84*, 35–42. doi: 10.1037/0022-0663.84.1.35

Marsh, H. W., & Craven, R. G. (1991). Self-other agreement on multiple dimensions of preadolescent self-concepts: Inferences by teachers, mothers, and fathers. *Journal of Educational Psychology, 83*, 393–404. doi:10.1037/0022-0663.83.3.393

Marsh, H. W., Smith, I. D., Barnes, J., & Butler, S. (1983). Self-concept: Reliability, stability, dimensionality, validity, and the measurement of change. *Journal of Educational Psychology, 75*, 772–790.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-

concept. *Review of Educational Research, 79*, 1129–1167. doi: 10.3102/0034654309337522

Pauli, C., Reusser, K., & Grob, U. (2007). Teaching for understanding and/or self-regulated learning? A video-based analysis of reform-oriented mathematics instruction in Switzerland. *International Journal of Educational Research, 46*, 294–305. doi:10.1016/j.ijer.2007.10.004

Pitman, E.J.G. (1939). A note on normal correlation. Biometrika, 31, 9–12.

Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging academic self-concepts of students. *Journal of Educational Research, 106*, 64–76. doi: 10.1080/00220671.2012.667010

Praetorius, A.-K., Greb, K., Lipowsky, F., & Gollwitzer, M. (2010). Lehrkräfte als Diagnostiker – Welche Rolle spielt die Schülerleistung bei der Einschätzung von mathematischen Selbstkonzepten? [Teachers' diagnostic perception – What role does students' grade performance play in the teacher's assessment of mathematical self-concepts?] *Journal für Bildungsforschung Online, 2*, 121–144.

Pretsch, J., Flunger, B., Heckmann, N., & Schmitt, M. (2013). Done in 60 seconds? Inferring teachers' subjective well-being from thin slices of nonverbal behavior. *Social Psychology of Education: An International Journal, 16*, 421–434. doi: 10.1007/s11218-013-9223-9

Schnotz, W., Horz, H., McElvany, N., Schroeder, S., Ullrich, M., Baumert, J., et al. (2010). Das BITE-Projekt: Integrative Verarbeitung von Texten und Bildern in der Sekundarstufe I [The Bite Project. Integrative processing of texts and images in lower secondary education]. In E. Klieme, D. Leutner, & M. Kenk (Eds.), Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes [Competence modelling. Interim evaluation of the DFG program and perspectives on the research approach]. *Zeitschrift für Pädagogik, 56* (pp. 143–153). Weinheim, Germany: Beltz.

Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2002). *SESSKO – Skalen zur Erfassung des schulischen Selbstkonzepts [Scales to assess the academic self-concept]*. Göttingen, Germany: Hogrefe.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie, 19*, 85–95. doi:10.1024/1010-0652.19.1.85

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*, 743–762.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 42*, 207–232.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78*, 546–558.

Zirkel, P., & Gable, R. K. (1977). The reliability and validity of various measures of self-concept among ethnically different adolescents. *Measurement & Evaluation in Guidance, 10*, 48–54.