Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Teacher Humour in the Classroom: Comparing Teacher Reports, Student Reports, and Observer-Coded Humour Indicators

**Hassan Banaruee**

hassan.banaruee@ph-weingarten.de

University of Education Weingarten

**Julian Wagner**

University of Freiburg

**Sonja Bieg**

University of Education Weingarten

**Research Article**

**Keywords:** Teacher humour, Teacher report, Student report, Observer ratings

**Posted Date:** June 30th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9969185/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.

## Abstract

Since teacher humour is relational, context-dependent, and dependent on interpretation, this study examined how four teacher humour types (course-related, course-unrelated, self-disparaging, and aggressive) are represented across teacher, student, and observer perspectives in authentic secondary school classrooms. Data were collected from 15 German classes, including 15 teachers, 376 students, and two trained observers per class. Teacher and student reports were assessed with parallel HUMLAS-based questionnaires, whereas observer ratings were collected using the HUMLAS-derived Coding System for Humour in the Classroom. Correlation analyses indicated perspective-dependent association patterns, with stronger convergence between teacher and student reports than between insider reports and observer ratings. Repeated-measures ANOVAs revealed significant perspective-related rating differences for all humour types. Teacher and student reports were generally higher than observer ratings for course-related and self-disparaging humour, whereas students reported more aggressive humour than teachers. Findings indicate systematic perceptual divergences in recognising teacher humour. The study's primary contribution is conceptual and methodological. It conceptualises teacher humour as a perspective-dependent classroom communication phenomenon and shows that teacher reports, student reports, and observer-coded indicators provide non-equivalent but complementary information about humour in classroom interaction. Empirically, the findings provide initial evidence that convergence and divergence differ by humour type and perceiver position.


## Introduction

Teacher humour is a relational classroom phenomenon rather than only an individual teacher behaviour. Teachers and students have shown to value humour as a feature of instructional quality (Banaruee & Bieg, 2026), even though they attribute different roles to teacher humour in the classroom, suggesting that their perspective might differ to some extent on this matter. Further, humour as a multi-dimensional construct is increasingly recognised as an effective instructional practice that shapes students' emotional, motivational, and cognitive experiences in the classroom (Bieg & Dresel, 2018; Bieg et al., 2026; Booth-Butterfield & Wanzer, 2016). Recent theoretical models of instructional humour emphasise that its educational impacts depend not only on teachers' humorous intentions but also on how humour is perceived by students (Bieg & Dresel, 2018; Wanzer et al., 2010). Accordingly, teacher humour should be examined from both teachers' and students' perspectives. Besides, external observation can provide a limited methodological perspective on humour indicators that are visible during classroom interaction.

Previous empirical research has relied predominantly on single-perspective self-reports, most often student ratings (e.g., Bieg & Dresel, 2018; Bishara, 2023; Bolkan et al., 2018; Liao et al., 2023; Wijaya et al., 2025) and less frequently teacher self-assessments (e.g., DiCamillo, 2010; Gerouki, 2011; Sahin, 2021). A limited number of studies have incorporated both teacher and student perspectives (Banaruee & Bieg, 2026; Bieg et al., 2026). Banaruee and Bieg (2026) qualitatively examined whether teachers and students conceptualise teacher humour as a feature of instructional quality. Their findings indicated convergence between teacher and student perspectives in identifying humour as a feature of instructional quality. Bieg et al. (2026), in contrast, used an experimental design across two parallel studies to examine how different types of teacher humour affect instructional quality, students' emotions, and motivation. Although the two studies used comparable methods and designs, teacher and student perspectives were not systematically compared. Together, these studies suggest that multiperspective evidence in teacher humour research remains limited, particularly with respect to direct comparisons between teachers, students, and external observers. Teacher and student self-reports provide valuable insights into humour as it is intended, experienced, and interpreted by classroom insiders. This assumption is consistent with classroom environment research, which distinguishes participant perceptions from external observation and argues that students and teachers may capture psychosocial aspects of classroom life that observers could miss or treat as less salient (Fraser, 1989). However,

insider reports do not address a different methodological question: which forms of humour can reliably be identified by trained external observers using a standardized coding framework.

In this respect, prior research argues that classroom observation offers a complementary methodological perspective with several well-established advantages (Kohake, 2024; Martinez et al., 2016). Adding observations to humour investigations enables research to compare insider interpretations with ratings produced through a predefined external coding framework (Helmke, 2014; Pianta & Hamre, 2009). Such comparisons can indicate which humour types are accessible across perspectives and which require relational or contextual knowledge. In addition, observer judgements are shaped by theoretical framework, attentional focus, and professional schemes, which may lead observers to attend to different cues than teachers and students (Mischo, 2012; Praetorius et al., 2012). Consequently, discrepancies between perspectives may reflect systematic perceptual and conceptual differences rooted in the way teacher humour is self-perceived, received, or observed. This issue is particularly relevant for teacher humour that often involves incongruity, ambiguity, and social interpretation (Frymier et al., 2008; Wanzer et al., 2010). Consequently, what teachers intend as humorous, what students experience as humorous, and what observers code as instructional humour may converge or diverge systematically. To date, no study has systematically examined teacher humour by integrating teacher self-reports, student ratings, and external observations within the same instructional context. As a result, instructional communication research lacks evidence about whether teacher humour is similarly represented when it is self-perceived by teachers, experienced by students, or identified by trained observers. This gap is not only methodological, but also conceptually important.

*1.1. Instructional Humour as a Relational Processing Phenomenon*

Instructional Humour Processing Theory (IHPT; Wanzer et al., 2010) provides the central theoretical basis for the present study. It offers a framework for understanding teacher humour processes, relying on incongruity-resolution theory (La Fave et al., 1996), disposition theory (Zillmann & Cantor, 1996), and elaboration likelihood model of persuasion (Petty & Cacioppo, 1986). IHPT posits that humour contains incongruent information that students must cognitively recognise and resolve. Only resolved and positively appraised humour leads to learning, whereas humour perceived as inappropriate can disrupt comprehension or strain teacher-student relationship (Wanzer & Frymier, 1999; Wanzer et al., 2010). Hence, IHPT conceptualises instructional humour not merely as a teacher behaviour, but as a process that depends on students' cognitive and evaluative processing. From this perspective, humour is not educationally meaningful simply because a teacher intends it to be humorous. It becomes instructionally relevant only when students identify, interpret, and appraise it within the ongoing classroom context. This assumption is central to the present study because it implies that teacher humour should not be assessed only from the teacher's perspective.

Although IHPT is the theoretical basis for the present study, it also has limitations. The theory was developed primarily to explain links between instructional humour and student learning, not to explain observer ratings or cross-perspective measurement discrepancies. To support relational processing, we draw on the conceptual model of instructional humour (Bieg & Dresel, 2018) and its developed questionnaire (HUMLAS; Bieg & Dresel, 2016). This conceptual model assumes different relational and causal connections between varying types of teacher humour and motivational, emotional, social, and cognitive dimensions of instructional quality. According to Bieg and Dresel (2018), affiliative teacher humour is positively associated with socio-emotional features, while aggressive teacher humour creates negative bonds, which is also supported by experimental research (Bieg et al., 2026). The meanings and effects of teacher humour depend on teachers' intentions and delivery, as well as students' interpretations within the classroom context (Frymier et al., 2008; Martin, 1996; Martin & Ford, 2018; Wanzer et al., 2010). This provides a rationale for examining teacher humour across multiple perspectives.

*1.2. From IHPT to Humour Types and the Conceptual Model*

The conceptual model of instructional humour by Bieg and Dresel (2018) extends the processing logic of IHPT by specifying how different humour types may relate to instructional quality, motivation, emotions, and classroom relationships. Whereas IHPT (Wanzer et al., 2010) explains why humour requires student processing and appraisal, the conceptual model (Bieg & Dresel, 2018) clearly differentiates the forms of teacher humour that may enter this process. This conceptual model identifies four types of teacher humour primarily based on students' perspectives (course-related, course-unrelated, self-disparaging, and aggressive). *Course-related humour*, which is the humour related to the teaching content, directly supports learning by embedding instructional content in humorous examples, thereby enhancing intrinsic motivation, positive emotions, and instructional quality (Bieg et al., 2026). *Course-unrelated humour*, in contrast, refers to the humour that is unrelated to the instructional content. *Self-disparaging humour* involves teachers' humorous situations targeting themselves. Research shows that it has acted heterogeneously in different studies (Wanzer et al., 2006). Recent experimental findings (Bieg et al., 2026) show that despite the pro-social essence of these three affiliative forms of humour, they differ in the way they affect students' perceived emotions, motivation, and instructional quality. *Aggressive humour*, on the contrary, is perceived as inappropriate since it targets and ridicules students. Although these humour forms are differentially associated with students' motivation, emotional engagement, and classroom climate (Bieg & Dresel, 2018; Stuart & Rosenfeld, 1994), it remains unclear to what extent these types are perceived similarly across teachers, students, and independent observers.

*1.3. Why Perspective May Converge or Diverge*

Since teacher humour is relational (Bieg & Dresel, 2018; Bieg et al., 2026), context-sensitive, and co-constructed in classroom interaction, no single perspective can be assumed to provide a complete account of its perceived functions and roles. Teacher reports, student reports, and observer ratings each may provide access to different perspectives of perceived humour. This distinction is consistent with Fraser's (1989) discussion of classroom environment assessment, in which external observation is differentiated from participant perceptions. Drawing on the concepts of alpha press and beta press (Murray, 1938), Fraser (1989) distinguished the classroom environment as assessed by detached observers from the environment as perceived by classroom participants. He further differentiated individual perceptions from shared class-level perceptions, a distinction that is directly relevant when comparing teacher reports, aggregated student reports, and observer ratings. From this perspective, cross-perspective discrepancies should not automatically be interpreted as measurement error, because each perspective may reflect a different standpoint on the same classroom phenomenon. Research argues that self-reports show strong validity in assessing internal states and deliberate instructional strategies relative to indirect measures such as response times or behavioural frequencies (Corneille & Gawronski, 2024).

Furthermore, a body of research has discussed the value of student ratings for assessing core dimensions of teaching quality and classroom interaction (Aleamoni, 1999; Lüdtke et al., 2006; Trautwein et al., 2006). In the case of humour, students' perceptions are particularly consequential, as humorous intent only becomes instructionally relevant if it is recognised and interpreted as such by students. Given that humour is not only delivered but also constructed through student reception, students' subjective interpretations may not necessarily align with teachers' intentions or with the frequency of actual humorous events.

On the other hand, observer ratings reflect the extent to which predefined humour categories are identifiable from an external theoretically guided standpoint by focusing on enacted instructional behaviour in real-time events (Helmke, 2014; Pianta & Hamre, 2009). At the same time, contemporary arguments in diagnostic theory emphasise that all classroom observation is inherently selective and theory-laden (Mischo, 2012). Observers may not capture the classroom event in its entirety but rather a defined set of behaviours guided and shaped by theoretical expectations

and the observational instruments. This limitation is especially salient for teacher humour, which can be subtle, situational, and dependent on contextual clues. As a result, humorous behaviour may be overlooked, differently weighted, or interpreted in ways that diverge from insider perspectives. This logic is consistent with multi-informant research showing that discrepancies between informants can contain meaningful information rather than merely reflecting measurement error (De Los Reyes & Kazdin, 2005). The value of a multiperspective approach therefore lies not in privileging one perspective over another, but in examining how these perspectives converge and diverge. Even though observer-based measures are well established in the study of instructional quality and classroom processes (Praetorius et al., 2014), no study to date has systematically incorporated systematic classroom observation into the investigation of teacher humour.

**The present study**

The present study examines how teacher humour is represented across teacher, student, and observer perspectives in authentic classroom communication. Consistent with classroom environment research that distinguishes participant perceptions from external observation, the present study compares teacher reports, aggregated student reports, and observer ratings as complementary but non-interchangeable perspectives on teacher humour (Fraser, 1989). Existing teacher humour research has relied mainly on single-perspective self-reports, particularly student ratings (e.g., Bieg & Dresel, 2018; Bishara, 2023; Wijaya et al., 2025). This has limited the field's capacity to determine whether teacher humour is a shared communicative phenomenon across classroom actors or whether different perspectives capture different aspects of it. The present study addresses this limitation by comparing teacher reports, aggregated student reports, and observer ratings within the same classrooms. The purpose is not to establish observer ratings as an objective criterion against which insider perceptions are evaluated, but to examine how each perspective makes different aspects of teacher humour visible. Cross-perspective convergence is therefore interpreted as evidence that a humour type is relatively accessible across perceiver positions, whereas divergence is interpreted as evidence that its identification may depend on classroom membership, relational positioning, temporal frame, or the operational definitions embedded in the observation framework.

Therefore, this study has two aims. The first aim is to examine whether the four theoretically defined humour types show different association patterns within teacher reports, student reports, and observer-coded humour indicators. Because the class-level sample is too small for confirmatory factor analyses across the three perspectives, within-perspective correlations are used to examine association patterns among the humour types. These correlations are not interpreted as evidence of discriminant validity. Additional scale-level evidence is provided through internal consistency estimates and, for student reports, aggregation indices that indicate whether student ratings can be meaningfully represented at the class level.

The second aim is to examine convergence and divergence across perspectives. Teacher and student reports are treated as insider perspectives with access to broader classroom experience, whereas observer ratings are treated as a bounded external perspective on humour indicators visible during one observed lesson. These three perspectives are not treated as interchangeable measures of the same construct. Rather, they are used to examine whether theoretically defined humour types show similar or different representation patterns across perspectives. Accordingly, we formulated the following hypotheses:

H1: The four humour types will show different within-perspective association patterns across teacher reports, student reports, and observer-coded humour indicators.

Affiliative humour types are generally characterised by a pro-social intent, lower interpersonal risk, and greater normative acceptance than aggressive humour (Wanzer et al., 2010). Because they are more likely to be interpreted

consistently, affiliative humour forms should show greater agreement across perspectives than aggressive humour.

H2: Teacher and student reports will show stronger convergence for affiliative humour types than for aggressive humour.

Teacher and student reports may also differ for affiliative humour types because teachers have direct access to their own instructional intentions and may recognise humour attempts that are not necessarily experienced by students as humorous. Course-related and self-disparaging humour are generally consistent with constructive and affiliative uses of humour (Bieg & Dresel, 2018; Wanzer et al., 2010), and teachers may therefore report these behaviours as part of their intended instructional practice. Students, in contrast, report humour from the position of recipients and may endorse these humour types only when the humour is recognised, interpreted, and remembered as such. Therefore, teachers are expected to report higher levels of course-related and self-disparaging humour than students.

H3: Teachers will report higher levels of course-related and self-disparaging humour than students.

Observer ratings rely on overt behavioural cues and predefined coding criteria, which may make less instructional and more salient humorous events particularly noticeable from an external perspective. Course-unrelated humour, which often involves conspicuous deviations from instructional flow, may therefore be emphasised more strongly by observers than by classroom insiders.

H4: Observers will rate course-unrelated humour higher than teachers and students.

Aggressive humour is often subtle, relationally embedded, and directed at students, who are the immediate recipients and evaluators of its interpersonal impact (Bieg et al., 2026; Frymier et al., 2008; Wanzer et al., 2006). Teachers may underreport such humour due to social desirability, and observers may detect only its externally visible indicators.

H5: Students will report higher levels of aggressive humour than teachers and observers because they are the immediate recipients of its meaning, while observer ratings of aggressive humour will be interpreted as indicators of externally visible aggressive humour during the observed lesson.

# Methods

### Participants

The study involved 15 secondary school teachers and their corresponding 15 ninth-grade classes from Germany. All participating teachers were officially recruited secondary school English teachers, resulting in a homogenous instructional context with respect to subject matter and curricular demands. Teachers' ages ranged from 29 to 53 years ($M$ = 39.9, $SD$ = 7.8), representing a predominantly mid-career sample. Teaching experience varied between 1 and 18 years ($M$ = 10.9, $SD$ = 5.8). The majority were female (n = 12), with three male teachers.

Student data were collected from 376 ninth-grade students aged between 13 to 16 years ($M$ = 14.2, $SD$ = 0.5), with the majority being 14 years old (74%). Gender information indicated that 51.2% of students identified as female, 45.2% as male, and 3.6% as diverse or non-binary. The focus on ninth-grade students is theoretically and methodologically aligned. At this educational stage, students typically possess sufficient cognitive, linguistic, and socio-emotional maturity to recognise, interpret, and evaluate nuanced forms of teacher humour, including irony, self-disparagement, and aggressive humour. Moreover, ninth grade is also the focal grade level of the German PISA-Ceco study, which investigates classroom processes using proximal, multiperspective assessments (teacher and student questionnaires as well as structured lesson observations; Todtenhöfer et al., 2025).

Observer data were provided by two trained observers who were not involved in teaching the participating classes and had no prior relationship with the participating teachers or students. The first observer was a researcher with a PhD in English didactics and teaching and served as the lead researcher in the broader longitudinal teacher humour project. The second observer was a research assistant with a bachelor's degree in psychology who had been involved in the project for two years, primarily in data collection. Both observers received standardized training in the observation framework before coding. The training was conducted by a supervising researcher who contributed to the development of the HUMLAS and Coding System for Humour in the Classroom (CSHC) coding frameworks. All participating students provided informed assent and written parental consent since they were minors. The study protocol was approved by the University of Education XX Research Ethics Committee, and was preregistered as part of the funding provided by XX.

**Procedure**

Teachers and students completed self-report questionnaires assessing teacher humour using HUMLAS (Bieg & Dresel, 2016). While teachers completed paper-pencil questionnaires, students' questionnaires were administered online in schools' computer rooms. Both teachers and students in one class completed the questionnaires at the same time. Corresponding classroom observations were conducted one or two weeks before the administration of teacher and student questionnaires. This ensured temporal proximity between reported and observed teacher humour, allowing cross-perspective comparisons.

CSHC was developed to adapt the HUMLAS questionnaire to structured classroom observation. The items were parallel to the HUMLAS items (Bieg & Dresel, 2016) for each type of humour. Each of the 15 classes was observed once during a regular 45-minute lesson. Real-time independent coding was used because the study aimed to examine whether predefined humour categories could be identified during ordinary classroom interaction without relying on retrospective reconstruction. This procedure preserved the temporal flow of the lesson and allowed observers to register humour as it occurred. Praetorius et al. (2014) found classroom management and student support stable dimensions of instructional quality and only one lesson of observation per teacher supports scientific measurement of the observed phenomenon in the actual classroom. Since teacher humour is associated with student support (Banaruee & Bieg, 2026; Bieg et al., 2026; Stuart & Rosenfeld, 1994; Wanzer et al., 2010), it is reasonable to accept that one observation session per teacher provides information to assess the observers' perspective.

Classroom observations were conducted by two trained observers per lesson. Both observers attended the same real-time classroom session simultaneously, with one observer seated on the left side of the classroom and the other on the right side. Observers did not communicate during the lesson or after the observation before submitting their ratings. Before data collection, observers completed structured training that included familiarisation with the theoretical definitions of humour types, discussion of item-level coding rules, practice coding with non-sample classroom videos, and calibration sessions in which discrepancies were discussed to establish a shared understanding of the coding categories. During training, observers were instructed to code a behaviour only when it contained an identifiable humour cue, such as incongruity, joking tone, humorous facial expression, or an explicitly funny anecdote. Smiling, general friendliness, or ordinary classroom talk were not coded unless connected to a recognisable humorous cue. Sarcasm and irony were coded only when the humorous frame was sufficiently identifiable from the teacher's wording, tone, or delivery. Student-directed remarks were coded as aggressive humour only when the student was the target of ridicule, or teasing related to ability or personal characteristics.

Both raters were blind to teachers' and students' self-report data. This procedure was used to reduce potential bias and to account for rater-specific variance in observational judgments (Praetorius et al., 2012). The use of two raters per lesson also allowed interrater reliability to be estimated. The observation protocol was intentionally focused on 17

theoretically anchored humour items rather than broad, multidimensional instructional constructs, thereby reducing the complexity of the rating task. This focused design is consistent with rater research showing that complex assessment tasks can increase cognitive demands and reduce rating accuracy (Gaugler & Thornton, 1989; Lievens, 1998). By restricting the coding domain to teacher humour and using a theoretically defined coding framework, the present study differs from prior observational work in which humour was examined in broader classroom contexts without a comparably focused observation scheme (e.g., Miller et al., 2017; Miura & Jones, 2005).

**Instruments**

Teacher humour was assessed using teacher self-reports, student reports, and structured classroom observations. The full wording of all questionnaire and observation items, their assignment to the four humour dimensions, and the response or coding formats are provided in Appendices A and B. Appendix A reports the HUMLAS items used for teacher and student self-reports, including the teacher-adapted wording. Appendix B reports the complete CSHC observation checklist, including the item-level behavioural indicators used by trained observers.

Teacher-reported humour was assessed using a teacher-adapted version of the Teacher Humour from Students' Perspective Scale (HUMLAS; Bieg & Dresel, 2016). The *teacher* version comprises 17 items rated on a five-point Likert scale ranging from 1 (never) to 5 (very often). All items are introduced by the question, "How do you conduct English lessons in this class?", and the prompt "I …". The scale includes four subscales: course-related humour (6 items; α = .91; e.g., "… use humour in relation to the teaching material"), course-unrelated humour (3 items; α = .86; e.g., "… tell jokes that have nothing to do with the course material"), self-disparaging humour (4 items; α = .82; e.g., "… make fun of myself in class"), and aggressive humour (3 items; α = .61; e.g., "… make fun of students in class"). One aggressive humour item was excluded from the reliability estimation because it showed zero variance in the present sample. The full teacher-adapted item wording is presented in Appendix A1.

*Student* perceptions of teacher humour were assessed using the student version of HUMLAS (Bieg & Dresel, 2016). This version includes the same abovementioned four humour dimensions as the teacher version, with items phrased from the students' perspective. The subscales were course-related humour (6 items; α = .94; e.g., "Our teacher uses explanations and illustrations that are interesting or fun during the course"), course-unrelated humour (3 items; α = .88; e.g., "Our teacher talks about funny topics that are unrelated to the course"), self-disparaging humour (4 items; α = .86; e.g., "Our teacher turns himself/herself into an object of ridicule after making mistakes during the course"), and aggressive humour (4 items; α = .84; e.g., "Our teacher ridicules students' intellectual capacity during the course"). The complete student-report item wording is provided in Appendix A2.

*Observed* teacher humour was coded using the Coding System for Humour in the Classroom (CSHC), a HUMLAS-derived observational adaptation designed to capture externally visible indicators of teacher humour during live classroom instruction. The CSHC consisted of 17 items: six items for course-related humour (α = .96; e.g., "The teacher uses a humorous example to explain or illustrate something."), three for course-unrelated humour (α = .66; e.g., "The teacher tells a funny anecdote that is not related to the lesson content."), four for self-disparaging humour (α = .76; e.g., "The teacher tells a funny anecdote or story about themselves that is embarrassing."), and four for aggressive humour (α = .86; e.g., "The teacher jokes about a student's intelligence."). The complete checklist, including all item-level behavioural indicators and their assignment to humour dimensions, is presented in Appendix B.

The CSHC was developed as a deductive observational adaptation of the HUMLAS questionnaire. Each HUMLAS item was translated into a visible or audible classroom indicator that could be identified by trained observers during live instruction. The coding system therefore focused on teacher-produced humour indicators, such as humorous

examples, jokes, funny anecdotes, self-directed humorous comments, exaggerated vocal expression, gestures, and student-directed teasing. Items that required direct access to teacher intention or student appraisal were reformulated as observable behaviours. Thus, CSHC ratings should be understood as observer-coded indicators of externally visible humour attempts during one lesson, not as measures of students' humour appreciation, cognitive resolution, or emotional appraisal. For each observed humour event, raters recorded both the humour category and the frequency of occurrence. Observed frequencies were subsequently transformed into a five-point ordinal scale to align the observer data with the response format of the teacher and student self-report measures: 1 = no occurrence, 2 = once, 3 = two to three times, 4 = four to five times, and 5 = more than five times.

**Data Analyses**

Analyses followed Classical Test Theory principles (Moosbrugger, 2012). Item-level statistics and internal consistency estimates were computed for all questionnaire scales and observational ratings. Internal consistency was estimated using Cronbach's α. Intraclass correlation coefficients (ICCs) were used for two distinct purposes: first, to evaluate whether student ratings could be aggregated to the class level, and second, to assess interrater reliability for the observer ratings. These two applications of ICCs address different methodological questions and are therefore reported separately.

Because student ratings were collected at the individual level, whereas teacher and observer ratings were available at the class level, student ratings were aggregated to class means before the main analyses to ensure comparability across perspectives. Following Lüdtke et al. (2006), the appropriateness of aggregation was evaluated using ICC(1) and ICC(2), which assess the extent of between-class variance in individual student ratings and the reliability of the aggregated class-level means, respectively. ICC(1) reflects the proportion of total variance in student ratings that is attributable to between-class differences. It therefore indicates the extent to which students within the same class provide ratings that are more similar to one another than ratings provided by students in different classes. ICC(2), in contrast, estimates the reliability of the aggregated class mean and increases as the number of student raters per class increases (Bliese, 2000; Lüdtke et al., 2006). ICC(1) values were .28 for course-related humour, .12 for course-unrelated humour, .22 for self-disparaging humour, and .11 for aggressive humour, indicating systematic between-class variance in students' ratings. The corresponding ICC(2) values were .91, .77, .87, and .76, respectively, indicating acceptable to high reliability of the class-level means. These results support the aggregation of student ratings to the class level. Accordingly, class-level student means were used in the subsequent analyses.

For the observational data, interrater reliability was evaluated using two-way mixed-effects ICCs with a consistency definition, because the same two observers rated each lesson and the primary aim was to assess the consistency of their ratings rather than absolute agreement. ICCs were .96 for course-related humour, .66 for course-unrelated humour, .76 for self-disparaging humour, and .86 for aggressive humour, indicating acceptable to excellent agreement according to Cicchetti's (1994) criteria. Because the observer ratings showed sufficient interrater reliability overall, the two raters' scores were averaged for each humour type to reduce rater-specific variance. The averaged observer scores were used in all subsequent analyses.

To examine the empirical distinctness and convergence of humour types across perspectives, bivariate correlations were calculated among teacher-, student-, and observer-reported humour dimensions, corresponding to H1 and H2. Correlations were appropriate because these hypotheses concerned the degree of association among humour dimensions within and across perspectives rather than differences in mean levels. To test perspective differences in mean levels of teacher humour, corresponding to H3, H4, and H5, repeated-measures analyses of variance were conducted. Repeated-measures ANOVAs were appropriate because the same classrooms had ratings from all three perspectives, allowing mean differences among teacher, student, and observer ratings to be examined within the

same class-level units. These analyses were interpreted cautiously as exploratory tests of perspective-dependent rating patterns rather than as evidence that any perspective overestimated or underestimated teacher humour relative to another. All statistical tests were two-tailed with an alpha level of $p < .05$. All analyses were conducted using SPSS Version 30.

## Results

Descriptive statistics and bivariate correlations among humour variables from all perspectives are presented in Table 1.

Hypothesis 1 stated that the four humour types would show different within-perspective association patterns across teacher reports, student reports, and observer-coded humour indicators. To test this hypothesis, within-perspective correlations were examined separately for teacher reports, aggregated student reports, and observer ratings. As shown in Table 1, the mean scores indicated that affiliative humour types were generally rated higher than aggressive humour across perspectives. The only exception was found in observer ratings, where aggressive humour was marginally higher than course-unrelated humour, although this difference was small. Among teachers, course-related humour was positively associated with course-unrelated humour, $r = .61$, $p < .05$; course-unrelated humour was positively associated with self-disparaging humour, $r = .67$, $p < .01$; and aggressive humour was positively associated with self-disparaging humour, $r = .52$, $p < .05$. Among students, course-related humour was positively associated with self-disparaging humour, $r = .60$, $p < .05$. In contrast, no significant associations among the four humour types were found in observer ratings.

Second, to address H2, we examine cross-perspective convergence between teacher reports, aggregated student reports, and observer ratings, with particular attention to corresponding humour types across perspectives. A significant corresponding association emerged between teacher and student reports for course-related humour, $r = .60$, $p < .05$. Student-reported course-related humour was also positively associated with teacher-reported course-unrelated humour, $r = .68$, $p < .01$, and student-reported self-disparaging humour was positively associated with teacher-reported course-unrelated humour, $r = .62$, $p < .05$. In contrast, observer ratings showed limited convergence with teacher and student reports. The only significant association involving observer ratings was a negative correlation between teacher-reported aggressive humour and observer-rated course-unrelated humour, $r = -.52$, $p < .05$. Aggressive humour showed no significant convergence across corresponding teacher, student, and observer perspectives.

**Table 1** *Descriptive statistics, and Correlations for Study Variables*

| | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Teacher reports* | | | | | | | | | | | | | | |
| 1. Course-related | 3.30 | 0.82 | _ | | | | | | | | | | | |
| 2. Course-unrelated | 1.91 | 1.07 | .61* | _ | | | | | | | | | | |
| 3. Self-disparaging | 2.62 | 0.69 | .46 | .67** | _ | | | | | | | | | |
| 4. Aggressive | 1.17 | 0.15 | .12 | .42 | .52* | _ | | | | | | | | |
| *Student aggregated evaluations* | | | | | | | | | | | | | | |
| 5. Course-related | 2.95 | 0.63 | .60* | .68** | .21 | .07 | _ | | | | | | | |
| 6. Course-unrelated | 1.81 | 0.42 | .12 | -.03 | -.22 | -.26 | .41 | _ | | | | | | |
| 7. Self-disparaging | 2.42 | 0.50 | .16 | .62* | .43 | .25 | .60* | .34 | _ | | | | | |
| 8. Aggressive | 1.53 | 0.44 | -.17 | -.23 | .18 | -.11 | -.25 | .43 | .07 | _ | | | | |
| *Observations* | | | | | | | | | | | | | | |
| 9. Course-related | 1.53 | 0.32 | .15 | -.08 | -.24 | -.40 | .47 | .18 | -.03 | -.33 | _ | | | |
| 10. Course-unrelated | 1.12 | 0.16 | -.01 | -.39 | -.06 | -.52* | -.24 | .37 | -.04 | .44 | .01 | _ | | |
| 11. Self-disparaging | 1.32 | 0.26 | .21 | -.24 | .13 | -.36 | -.07 | -.21 | -.37 | .13 | .33 | .33 | _ | |
| 12. Aggressive | 1.28 | 0.25 | .35 | .00 | .20 | .08 | .27 | .17 | .25 | .24 | .15 | .15 | .49 | _ |

*Note.* N = 15 classes. *$p < .05$ **$p < .01$. (two-tailed).

**Table 2** *Multivariate Test for the Effect of Perspective on Types of Teacher Humour*

| Effect | Wilks' $\Lambda$ | *F* | df | *p* | $\eta p^2$ |
|---|---|---|---|---|---|
| Course-related humour | .117 | 48.99 | 2, 13 | < .001 | .883 |
| Course-unrelated humour | .217 | 23.41 | 2, 13 | < .001 | .783 |
| Self-disparaging humour | .182 | 29.29 | 2, 13 | < .001 | .818 |
| Aggressive humour | .610 | 4.16 | 2, 13 | .04 | .390 |

Third, to address H3 to H5, we report perspective-related rating patterns using repeated-measures analyses and Bonferroni-adjusted pairwise comparisons. Table 2 presents the omnibus repeated-measures effects of perspective

for each humour type. These analyses were conducted before the hypothesis-specific pairwise comparisons relevant to H3 to H5, which showed that the effect of perspective was significant for all four humour types. The largest effect emerged for course-related humour, Wilks' $\Lambda$ = .117, $F(2, 13) = 48.99$, $p < .001$, $\eta p^2 = .883$. Large effects were also found for self-disparaging humour, Wilks' $\Lambda$ = .182, $F(2, 13) = 29.29$, $p < .001$, $\eta p^2 = .818$, and course-unrelated humour, Wilks' $\Lambda$ = .217, $F(2, 13) = 23.41$, $p < .001$, $\eta p^2 = .783$. The effect for aggressive humour was smaller but statistically significant, Wilks' $\Lambda$ = .610, $F(2, 13) = 4.16$, $p = .040$, $\eta p^2 = .390$. Thus, the overall perspective effect was strongest for course-related humour and weakest for aggressive humour.

Table 3 presents the Bonferroni-adjusted pairwise comparisons used to evaluate the direction and location of these perspective effects. For course-related humour, teacher and student ratings did not differ significantly. However, both teacher ratings and student ratings were significantly higher than observer ratings, with the largest difference observed between teachers and observers, *M*diff = 1.77, $p < .001$, followed by students and observers, *M*diff = 1.42, $p < .001$. For course-unrelated humour, the only significant pairwise difference was between students and observers, with students reporting higher levels than observers rated, *M*diff = 0.69, $p < .001$. The teacher-observer difference did not reach statistical significance, and teacher and student reports did not differ significantly. For self-disparaging humour, teacher and student ratings did not differ significantly. However, observer ratings were significantly lower than both teacher ratings, *M*diff = 1.30, $p < .001$, and student ratings, *M*diff = 1.10, $p < .001$. For aggressive humour, students reported significantly higher levels than teachers, *M*diff = 0.37, $p = .031$, whereas observer ratings did not differ significantly from teacher ratings or student ratings.

**Table 3** *Paired Comparisons of Perceived Teacher Humour across Teacher, Student, and Observer Perspectives*

| Humour type | Compared perspectives | *M*diff | SE | *p* | 95% CI |
|---|---|---|---|---|---|
| Course-related | T – S | 0.35 | 0.17 | .182 | –0.12, 0.82 |
| | T – O | 1.77* | 0.22 | < .001 | 1.19, 2.36 |
| | S – O | 1.42* | 0.14 | < .001 | 1.03, 1.81 |
| Course-unrelated | T – S | 0.10 | 0.30 | 1.00 | –0.71, 0.92 |
| | T – O | 0.79 | 0.30 | .055 | –0.02, 1.59 |
| | S – O | 0.69* | 0.10 | < .001 | 0.41, 0.96 |
| Self-disparaging | T – S | 0.20 | 0.17 | .763 | –0.26, 0.66 |
| | T – O | 1.30* | 0.18 | < .001 | 0.81, 1.79 |
| | S – O | 1.10* | 0.17 | < .001 | 0.65, 1.55 |
| Aggressive | T – S | –0.37* | 0.12 | .031 | –0.71, –0.03 |
| | T – O | –0.12 | 0.07 | .392 | –0.31, 0.08 |
| | S – O | 0.25 | 0.12 | .144 | –0.06, 0.56 |

*Note.* T = Teacher, S = Student, O = Observer. CI = confidence interval. *p* values are Bonferroni-adjusted.

Taken together, the results provided partial support for the hypotheses. Hypothesis 1 was supported, as the four humour types showed differentiated within-perspective association patterns. Several significant associations were found within teacher reports, one significant association was found within student reports, and no significant associations were found within observer ratings. Hypothesis 2 was partially supported. Teacher and student reports

converged significantly for course-related humour, and aggressive humour showed no significant convergence across corresponding perspectives. However, convergence was not consistent across all affiliative humour types, and observer ratings showed limited convergence with teacher and student reports. Hypothesis 3 was not supported. Although teachers reported descriptively higher levels of course-related and self-disparaging humour than students, these differences were not statistically significant. Hypothesis 4 was again partially supported. Observer ratings of course-unrelated humour differed significantly from student reports, but not from teacher reports. However, the significant difference was in the opposite direction than hypothesised since observers rated course-unrelated humour lower than students, rather than higher. Hypothesis 5 was partially supported. Students reported significantly higher levels of aggressive humour than teachers, but student reports did not differ significantly from observer ratings.

## Discussion

The aim of the present study was to examine how theoretically defined teacher humour types are represented across teacher reports, student reports, and observer ratings in authentic secondary school classrooms. The findings showed three main patterns. First, the four humour types showed perspective-dependent association patterns, with clearer interrelations in teacher reports and more limited interrelations in student reports, whereas observer ratings showed no significant associations among humour types. Second, convergence was strongest between teacher and student reports and weaker when observer ratings were involved. Third, teacher and student reports showed higher values than observer ratings for course-related and self-disparaging humour, whereas for course-unrelated humour only students reported significantly higher levels than observers. For aggressive humour, students reported higher levels than teachers.

The observed convergence and divergence across teacher reports, aggregated student reports, and observer ratings support the assumption that these perspectives provide complementary but non-interchangeable information about teacher humour. This interpretation is consistent with classroom environment research, which distinguishes participant perceptions from external observation and cautions against treating one perspective as a simple substitute for another (Fraser, 1989). Rather than indicating only measurement inconsistency, discrepancies across perspectives may reflect differences in whether teacher humour is self-perceived by teachers, experienced by students, or identified by observers through visible behavioural indicators.

Our results are consistent with the theoretical premise that instructional humour depends on recognition, appraisal, and interpretation (Wanzer et al., 2010), while also supporting the usefulness of differentiating humour types as proposed in the conceptual model of instructional humour (Bieg & Dresel, 2018). Consistent with prior research (Bieg & Dresel, 2018; Bieg et al., 2026; Frymier et al., 2008), the four humour types showed distinguishable patterns. The correlational patterns varied across perspectives: among teachers, course-related humour correlated with course-unrelated humour, and self-disparaging humour correlated with both course-unrelated and aggressive humour. Among students, course-related and self-disparaging humour were moderately associated. These findings extend earlier self-report studies (Bieg et al., 2019, 2026; Wanzer et al., 2010) by showing that functional distinctions among humour types are more clearly visible in insider reports than in observer ratings of real-time classroom behaviour. The absence of significant associations among observer-rated humour types suggests that external observations may capture a more restricted and temporally bounded representation of teacher humour.

The convergence pattern suggests that affiliative humour was more consistently shared between teachers and students than between insider and observer perspectives. Teacher and student reports showed significant convergence only for course-related humour. For course-unrelated and self-disparaging humour, teacher and student reports did not significantly correlate. These findings indicate that teacher-student convergence was limited to the

affiliative humour type most directly embedded in instructional content. However, observers' ratings did not align with either teachers or students for any affiliative type. This discrepancy may indicate that affiliative humour is less salient to external observers, particularly because observers based their ratings on a single lesson, whereas teachers and students likely drew on experiences across the school year. In contrast, aggressive humour showed no significant convergence across corresponding perspectives, and students reported significantly more aggressive humour than teachers. However, observer ratings did not differ significantly from either teacher or student ratings. This pattern suggests that aggressive humour may be especially sensitive to relational positioning, while also indicating that observer ratings should not be interpreted as simply lower or higher than insider reports. These findings are consistent with Disposition Theory (Zillmann & Cantor, 1972, 1996), which emphasises that humour evaluation depends on social alignment with the humour target. Students are less likely to perceive humour as appropriate when directed at themselves or their in-group (Zillmann & Cantor, 1972; La Fave et al., 1974). Thus, the limited convergence across humour types supports the value of a multiperspective approach.

Observer ratings were lower than teacher and student reports for course-related and self-disparaging humour. For course-unrelated humour, students reported significantly higher levels than observers, whereas the teacher-observer difference did not reach statistical significance. Thus, the clearest insider-observer differences were found for course-related and self-disparaging humour. While teachers perceive themselves as using self-disparaging humour in the classroom, such instances appear less perceived by observers as outsiders. This pattern suggests that external observers may represent affiliative humour differently from classroom insiders, particularly when humour depends on shared classroom history, prior interactions, or implicit relational cues. The difference between student reports and observer ratings for course-unrelated humour may indicate that this humour type is more situational and depends on specific opportunities within the lesson. In the observed lessons, such opportunities may not have occurred or may have been less visible to external observers, whereas students may have interpreted some moments within the lived instructional flow as course-unrelated humour.

Teacher and student reports did not differ significantly for any affiliative humour type. This pattern indicates that teachers and students reported comparable average levels of course-related, course-unrelated, and self-disparaging humour. One possible interpretation is that both groups draw on a shared classroom history when evaluating affiliative humour, which may support similar general perceptions of its relational meaning. However, this interpretation should be limited to mean-level similarity. The correlational findings indicated significant teacher-student convergence only for course-related humour. Thus, teachers and students reported similar average levels of affiliative humour, but they did not consistently agree on the relative occurrence of all affiliative humour types across classrooms. This distinction suggests that shared classroom membership may contribute to comparable overall evaluations without producing consistent teacher-student agreement for each affiliative humour type. This difference should therefore not be interpreted as simple teacher overestimation.

Teachers' lower self-ratings of aggressive humour relative to students may reflect reduced awareness of subtle paralinguistic cues or social desirability biases (Dunning et al., 2004; Podsakoff et al., 2003). Because aggressive humour often involves tone, sarcasm, or facial expressions, teachers may overlook how their remarks are interpreted by students. Observers, who lacked an ongoing relational history with the class, may have interpreted aggressive humour cues differently from both teachers and students (Good & Lavigne, 2017; Pianta, 1999). These findings emphasise the limitations of relying solely on single-perspective self-reports and illustrate the diagnostic value of incorporating multiple perspectives to detect potentially harmful teacher humour. In the framework of multiperspective assessment (De Los Reyes & Kazdin, 2005), such patterns represent divergent yet valid reports, reflecting differential access to behavioral, cognitive, and affective components of classroom humour.

# Theoretical and Practical Implications

Our findings contribute to instructional communication research by clarifying how teacher humour varies across perceiver positions. Rather than extending IHPT through a direct test of its assumptions about learning, the present study uses IHPT to specify why humour should be examined as a reception-dependent classroom phenomenon. IHPT emphasises that instructional humour becomes educationally meaningful only when students recognise, process, and appraise it (Wanzer et al., 2010). The present findings add that this processing logic also has measurement implications: teacher humour may be represented differently across teacher, student, and observer data. Accordingly, the contribution is not that one data source provides the most accurate estimate, but that each indexes a different position in the humour process.

The methodological contribution is also specific. This study highlights the value of incorporating multiple perspectives into research on teacher humour as a feature of teaching quality (Banaruee & Bieg, 2026). Although observer ratings have long been used to assess instructional practices (Praetorius et al., 2014), they have rarely been applied to teacher humour as a distinct facet of classroom interaction. The present findings show that structured observation can function as a bounded external lens on classroom humour. Observer ratings did not reproduce teacher and student reports, especially for humour types that may require additional classroom knowledge. This divergence is informative because it indicates that humour in teaching cannot be understood only as the frequency of humorous events. It must also be understood in relation to the information available to each data source. These discrepancies should not be interpreted as a weakness of insider perspectives or as evidence that observer ratings are superior. Instead, they show that the CSHC captures the external visibility of predefined humour indicators. The findings therefore provide an empirical basis for refining the observation framework. In particular, categories with low convergence between observer and insider reports may require more explicit behavioural anchors, examples of difficult-to-observe humour, or supplementary coding notes that help observers distinguish absence of humour from humour that is not externally salient.

Our findings have clear implications for teacher education regarding the intentional use of different types of teacher humour. Teacher education should address possible teacher-student discrepancies in humour perception. This is particularly relevant for aggressive humour, where students reported higher levels than teachers. Reflection on humour use may therefore help teachers evaluate how humorous remarks are received by students. Because the present study did not assess student outcomes, these implications should be understood as implications for reflective practice rather than as evidence for specific intervention effects. Encouraging teachers to reflect on course-related humour and potentially sensitive forms of humour, particularly aggressive humour, may support more deliberate humour use in classroom interaction (Bieg & Dresel, 2018; Bieg et al., 2026; Stuart & Rosenfeld, 1994; Wanzer et al., 2010).

# Limitations and Future Directions

Despite these contributions, several limitations should be considered when interpreting the findings. The first limitation concerns the use of a single observed lesson and the different reference frames of the measures. A single 45-minute observation cannot provide a stable trait-level estimate of a teacher's general humour use or support strong conclusions about typical humour frequency across the school year. This design is suitable for examining whether predefined humour categories can be coded reliably in authentic classroom interaction, but observer ratings captured one specific lesson, whereas teacher and student questionnaires assessed broader classroom experience. The five-point scales were therefore not metrically equivalent across perspectives, even though they were aligned formally. A student's endorsement of "very often" reflects accumulated experience across lessons, whereas an observer's highest category reflects more than five coded events in one lesson. For this reason, comparisons

involving observer ratings should be interpreted as differences in external visibility within the observed lesson, not as direct differences in general humour frequency. Although Praetorius et al. (2014) showed that some dimensions of instructional quality can be reliably captured from a single lesson, teacher humour is likely more situation-specific and may fluctuate across temporal phases of the school year. Piwowar (2013) recommended multiple observations and/or video-based analyses to increase temporal stability. Future research would benefit from sampling several lessons per teacher, using repeated coding to capture the dynamic nature of humour over time, and collecting teacher and student reports during the same observed lessons to enable situational comparisons across all three perspectives.

Second, observer judgments are inherently selective and theory-laden. Following Mischo (2012), we assume that no observation is "holistic" or neutral; instead, diagnostic validity depends on making the focus of observation theoretically explicit and empirically regulated. In this study, we addressed this requirement by grounding the CSHC in HUMLAS (Bieg & Dresel, 2016, 2018) and by restricting the observational domain to clearly defined humour categories. While this narrowing of scope increases conceptual transparency and reduces information overload for raters (cf. Gaugler & Thornton, 1989; Lievens, 1998), it also limits the breadth of classroom phenomena captured and may overlook less frequent or more subtle forms of humour. Furthermore, observer ratings inevitably blend instructional signals with rater-specific variance, such as stringency, leniency, and different item interpretations (Praetorius et al., 2012). We mitigated these issues by using two trained raters per lesson and by examining interrater reliability at the item level, yet the findings should still be interpreted with these constraints in mind. Although the CSHC showed acceptable interrater reliability for most humour types, the present study provides only initial feasibility and reliability evidence for this HUMLAS-derived observation system. It does not establish the CSHC as a fully validated observational measure of teacher humour. Future studies should examine its construct validity, temporal stability across multiple lessons, sensitivity to low-frequency humour types, and convergence with event-based student appraisals.

Finally, teachers were aware that they were being observed, and some may therefore have altered their behaviour, potentially suppressing aggressive or course-unrelated humour. The extent of such reactivity cannot be determined from the present data, but it underscores the value of combining structured observation with student reports and, where feasible, less obtrusive approaches such as longer-term video-based observation. Although these limitations restrict the generalisability and temporal breadth of the conclusions, the study provides a methodological foundation for multiperspective research on classroom communication and highlights directions for future work on how perspectives on teacher humour converge and diverge within classroom context.

## Conclusion

This study provides initial evidence that teacher humour is not equally accessible across teacher, student, and observer perspectives. Significant perspective effects across all four humour types indicate that the three perspectives captured systematically different rating patterns. The study's contribution is primarily conceptual and methodological, with preliminary empirical evidence of perspective divergence. Conceptually, the findings support the view that humour in teaching is not only produced by teachers, but also interpreted by students and only partially identifiable through external observation. Methodologically, the study shows that comparing teacher, student, and observer perspectives can clarify which humour types are shared across perceiver positions and which are more dependent on relational history, contextual knowledge, or visible behavioural indicators. Empirically, the findings suggest, at the class level and within the limits of a single observed lesson, that affiliative humour showed greater teacher-student similarity at the mean level, whereas cross-perspective convergence was clearest for course-related humour. Aggressive humour, in contrast, showed greater teacher-student divergence. These conclusions should not

be interpreted as evidence that one perspective is superior to another. Teacher reports provide access to self-perceived humour practices, student reports provide access to humour as experienced by classroom participants, and observer ratings provide access to externally visible humour indicators. The main implication for instructional communication scholarship is that teacher humour should be understood as a communicative process whose assessment depends on who observes, experiences, or reports it.

## Declarations

# Author Contribution

HB: Conceptualisation; Methodology; Formal analysis; Investigation; Data curation; Writing – original draft, review & editing. JW: Investigation; Writing – review & editing. SB: Conceptualisation; Methodology; Investigation; Data curation; Supervision; Writing – review & editing.

# Acknowledgement

We admire the constant support from the schools' principals, teachers, and students in Baden-Württemberg, Saarland, and Thuringia in Germany throughout this research process. We acknowledge Deutsche Forschungsgemeinschaft for the funding and the established opportunity to contribute to this nuanced area in social sciences.

# Data Availability

Data Availability Statement: The data that support the findings of this study will be made openly available in the RADAR research data repository upon publication of the article.

## References

1. Aleamoni, L. M. (1999). Student Rating Myths Versus Research Facts from 1924 to 1998. *Journal of Personnel Evaluation in Education,* 13, 153–166. https://doi.org/10.1023/A:1008168421283
2. Banaruee, H. & Bieg, S. (2026). Is humour a feature of instructional quality?. *HUMOR*. https://doi.org/10.1515/humor-2025-0138
3. Bishara, S. (2023). Humor, motivation and achievements in mathematics in students with learning disabilities. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186X.2022.2162694
4. Bieg, S., & Dresel, M. (2016). Fragebogen zur Erfassung des Humors von Lehrkräften aus Schülersicht (HUMLAS): Konstruktion und Validierung [Construction and Validation of the German Questionnaire to Assess Students' Perceptions of Teacher Humor]. *Diagnostica*, *62*(1), 3–15. https://doi.org/10.1026/0012-1924/a000132
5. Bieg, S., & Dresel, M. (2018). Relevance of Perceived Teacher Humor Types for Instruction and Student Learning. *Social Psychology of Education, 21*(4)*,* 805–825. https://doi.org/10.1007/s11218-018-9428-z
6. Bieg, S., Grassinger, R., & Dresel, M. (2019). Teacher humor: Longitudinal effects on students' emotions. *European Journal of Psychology of Education, 34*(3), 517–534. https://doi.org/10.1007/s10212-018-0402-0
7. Bieg, S., Banaruee, H., & Dresel, M. (2026). The impact of teacher humour on teaching quality and student learning: An experimental approach. *Learning and Instruction, 102*, Article 102311. https://doi.org/10.1016/j.learninstruc.2025.102311

8. Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein, & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). Jossey-Bass.
9. Bolkan, S., Griffin, D. J., & Goodboy, A. K. (2018). Humor in the Classroom: The Effects of Integrated Humor on Student Learning. *Communication Education, 67*(2), 144–164. https://doi.org/10.1080/03634523.2017.1413199
10. Booth-Butterfield, M., & Wanzer, M. (2016). 9. Humor Enactment in Learning Environments. In P. Witt (Ed.), *Communication and Learning* (pp. 211–240). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9781501502446-010
11. Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284
12. Corneille, O., & Gawronski, B. (2024). Self-reports are better measurement instruments than implicit measures. *Nature Reviews Psychology, 3*, 835–846. https://doi.org/10.1038/s44159-024-00376-z
13. De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*(4), 483–509. https://doi.org/10.1037/0033-2909.131.4.483
14. DiCamillo, L. (2010). "Fun" Pedagogy Curtails Intellectual Rigor in a U.S. History Class. *The Journal of Social Studies Research*, *34*(2), 175–207. https://doi.org/10.1177/235227982010034002002
15. Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x
16. Fraser, B. J. (1989). Twenty years of classroom climate work: progress and Prospect. *Journal of Curriculum Studies, 21*(4), 307-327. https://doi.org/10.1080/0022027890210402
17. Frymier, A. B., Wanzer, M. B. & Wojtaszczyk, A. M. (2008). Assessing Students' Perceptions of Inappropriate and Appropriate Teacher Humor. *Communication Education, 57*(2), 266–288. https://doi.org/10.1080/03634520701687183
18. Gerouki, M. (2011). 'Teacher, will you marry me?' Integrating teachers' stories and self-reflective texts for exploring sexuality discourses in school settings. *Sex Education*, *11*(1), 1–12. https://doi.org/10.1080/14681811.2011.538140
19. Good, T. L., & Lavigne, A. L. (2017). *Looking in Classrooms* (11th ed.). Routledge. https://doi.org/10.4324/9781315627519
20. Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* [Instructional quality and teacher professionalism. Diagnosis, evaluation and improvement of instruction]. Klett-Kallmeyer.
21. Gaugler, B. B., & Thornton, G. C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology, 74*(4), 611–618. https://doi.org/10.1037/0021-9010.74.4.611
22. Kohake, K. (2024). Systematic observation to measure teaching quality in different contexts: Insights from science lessons, physical education lessons, and sports training using the Classroom Assessment Scoring System. *Frontiers in Education, 9*, Article 1437996. https://doi.org/10.3389/feduc.2024.1437996
23. La Fave, L., Haddad, J., & Maesen, W. A. (1996). Superiority, enhanced self-esteem, and perceived incongruity humour theory. In A. J. Chapman & H. C. Foot (Eds.), *Humor and laughter: Theory, research and applications* (pp. 63–91). New Brunswick, NJ: Transaction Publishers.

24. Liao, Y. H., Lee, M. F., Sung, Y. T., & Chen, H. C. (2023). The effects of humor intervention on teenagers' sense of humor, positive emotions, and learning ability: A positive psychological perspective. *Journal of Happiness Studies, 24*, 1–19. https://doi.org/10.1007/s10902-023-00654-2
25. Lievens, F. (1998). Factors which improve the construct validity of assessment centers: a review. *International Journal of Selection and Assessment, 6*, 141–152. https://doi.org/10.1111/1468-2389.00085
26. Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Reliability and agreement of student ratings of the classroom environment: A reanalysis of TIMSS data. *Learning Environments Research, 9*(3), 215–230. https://doi.org/10.1007/s10984-006-9014-8
27. Martin, R. A. (1996). The Situational Humor Response Questionnaire (SHRQ) and Coping Humor Scale (CHS): A decade of research findings. *HUMOR, 9*(3–4), 251–272. https://doi.org/10.1515/humr.1996.9.3-4.251
28. Martin, R. A., & Ford, T. E. (2018). *The psychology of humor: An integrative approach* (2nd ed.). Academic Press.
29. Martinez, F., Taut, S., & Schaaf, K. (2016). Classroom observation for evaluating and improving teaching: An international perspective. *Studies in Educational Evaluation, 49*, 15–29. https://doi.org/10.1016/j.stueduc.2016.03.002
30. Miller, J. L., Wilson, K., Miller, J., & Enomoto, K. (2017). Humorous materials to enhance active learning. *Higher Education Research & Development*, *36*(4), 791–806. https://doi.org/10.1080/07294360.2016.1238883
31. Mischo, C. (2012). Über den Umgang mit grundlegenden Problemen von Beobachtungs- und Diagnoseprozessen. *Frühe Bildung, 1*(2), 106-108. https://doi.org/10.1026/2191-9186/a000037
32. Miura, S., & Jones, R. (2005). LOL: The use of humor in secondary social studies classrooms. In L. P. McCoy (Ed.), *Studies in Teaching 2005 Research Digest* (pp. 71–75). Wake Forest University
33. Moosbrugger, H. (2012). Klassische Testtheorie (KTT). In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 103–117). Springer. https://doi.org/10.1007/978-3-642-20072-4_5
34. Murray, H. A. (1938). *Explorations in Personality*. Oxford University Press.
35. Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19,* 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2
36. Pianta, R. C. (1999). Why supportive relationships are essential. In R. C. Pianta, *Enhancing relationships between children and teachers* (pp. 7–23). American Psychological Association. https://doi.org/10.1037/10314-001
37. Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, *38*(2), 109-119. https://doi.org/10.3102/0013189X09332374
38. Piwowar, V. (2013). Einschätzung von Unterrichtsqualität durch externe Beobachterinnen und Beobachter. Eine kritische Betrachtung der aktuellen Vorgehensweise in der Schulpraxis. *Beiträge zur Lehrerbildung, 31*(2), 174-185. urn:nbn:de:0111-pedocs-138459
39. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
40. Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction, 22*(6), 387–400. https://doi.org/10.1016/j.learninstruc.2012.03.002
41. Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31,* 2–12. https://doi.org/10.1016/j.learninstruc.2013.12.002

42. Sahin, A. (2021). Humor Use in School Settings: The Perceptions of Teachers. *SAGE Open*, *11*(2), 1–16. https://doi.org/10.1177/21582440211022691
43. Stuart, W. D., & Rosenfeld, L. B. (1994). Student perceptions of teacher humor and classroom climate. *Communication Research Reports*, *11*(1), 87–97. https://doi.org/10.1080/08824099409359944
44. Swanson, P. B. (2013). Spanish teachers' sense of humor and student performance on the National Spanish Exams. *Foreign Language Annals, 46*(2), 146–156. https://doi.org/10.1111/flan.12031
45. Todtenhöfer, P., Heinle, A., Dümig, P., Tang, Y., Weis, M., Reinhold, F., Holzberger, D., Reiss, K., & Schiepe-Tiska, A. (2025). Fragebögen und Beobachtungsinstrumente in PISA-Ceco [Questionnaires and observation instruments in PISA-Ceco]. In A. Schiepe-Tiska, S. Krauss, K. Reiss, & D. Lewalter (Eds.), *PISA-Ceco – Dokumentation der Erhebungsinstrumente der Begleitstudie zu PISA 2022: Einblicke in den mathematisch-naturwissenschaftlichen Unterricht* (pp. 139–208). Waxmann. https://doi.org/10.31244/9783818850692
46. Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology, 90*(2), 334–349. https://doi.org/10.1037/0022-3514.90.2.334
47. Wanzer, M. B., & Frymier, A. B. (1999). The Relationship Between Student Perceptions of Instructor Humor and Students' Reports of Learning. *Communication Education, 48*(1), 48–62. https://doi.org/10.1080/03634529909379152
48. Wanzer, M. B., Frymier, A. B., Wojtaszczyk, A. M., & Smith, T. (2006). Appropriate and Inappropriate Uses of Humor by Teachers. *Communication Education, 55*(2), 178–196. https://doi.org/10.1080/03634520600566132
49. Wanzer, M. B., Frymier, A. B., & Irwin, J. (2010). An Explanation of the Relationship Between Instruction Humor and Student Learning: Instructional Humor Processing Theory. *Communication Education, 59*(1), 1–18. https://doi.org/10.1080/03634520903367238
50. Wijaya, T. T., Feng, M., Cao, Y., Su, M., & Habibi, A. (2025). Should mathematics teachers be humorous? A study based on the instructional humor processing theory. *Frontiers in Psychology, 16*, Article 1668887. https://doi.org/10.3389/fpsyg.2025.1668887
51. Zillmann, D., & Cantor, J. R. (1972). Directionality of transitory dominance as a communication variable affecting humour appreciation. *Journal of Personality and Social Psychology, 24*(2), 191–198. https://doi.org/10.1037/h0033384
52. Zillmann, D., & Cantor, J. R. (1996). A disposition theory of humour and mirth. In A. J. Chapman & H. C. Foot (Eds.), *Humor and laughter: Theory, research and applications* (pp. 93–115). New Brunswick, NJ: Transaction Publishers.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx