

A streamlined survey instrument to moderate university students' grades for group projects

University
students'
grades for
group projects

Stephen Dix

School of Management and Marketing, Curtin University, Perth, Australia

1767

Received 11 September 2023
Revised 4 January 2024
Accepted 5 January 2024

Abstract

Purpose – The aim of this paper is to generate a streamlined, transparent and effective instrument to fairly measure the contribution made by each student to a group project within a higher education context. The primary aim is to moderate the grades of underperforming students at the end of the project. There is a secondary benefit in alerting underperforming students to raise their contribution mid-task or face a potentially reduced grade at the final stage.

Design/methodology/approach – The development of this multi-dimensional instrument is guided by findings from previous research. The quest is to minimise the instructor's administrative work load in applying a moderation-only instrument that is open-source and available at no cost. Based on the literature, the survey instrument seeks to apply a peer-based, equitable and transparent evaluation of each member's contribution to a group task. The survey is applied at mid-task and again at end-task in order to afford underperformers the opportunity to address contribution deficits during the final phase of the project.

Findings – The instrument, called TANDEM©, offers a transparent, streamlined, equitable, confidential and practical measure of each student's contribution to a graded group task. Students whose end-task contribution falls below the group average rating receive a proportional reduction in their personal grade. Additionally, the end-task moderation instrument captures a single-item holistic measure of relative contribution that may, in the future, serve as a surrogate for the multi-dimensional measures currently in place.

Research limitations/implications – TANDEM© was developed with group sizes of four or five members in mind. There is no evidence to support its application to three-person groups. Moreover, the application was applied only amongst under-graduate students. It is yet to be applied across post-graduate groups and within online learning environments. Future research into diverse cultural settings would serve to advance understanding of how moderation is perceived across borders.

Practical implications – Several existing group grade moderation methods propose complex algorithms that are "black box" solutions from a student's perspective. In establishing a fair, streamlined, confidential and transparent process for peer-rated moderation, TANDEM© deploys a concise instrument with a relatively small administrative load. TANDEM © may be applied to all groups or can selectively be applied to groups that report moderate, strong or extreme levels of conflict.

Social implications – Students will appreciate the opportunity to rate peer contributions to group projects. This will dissipate the negative social sentiment that may arise when fellow students benefit from the work of others. Those students seeking conflict resolution within the group will value the transparent and equitable moderation of grades as well as the positive social implications that follow.

Originality/value – This research forms part of an ongoing quest to present a moderation instrument that fairly identifies student contribution to a group project. Whilst the solution proposed is one of many existing alternatives, its focus is on a practical moderation-only instrument that can immediately be applied to a course or major. The benefits lie in the ease of application and minimal administrative workload. This constitutes an original contribution to the individual (course or major) coordinator who seeks to apply a moderation-only instrument without having to commit to an extensive, broad-based group optimisation programme.

Keywords Moderation, Peer review, Tertiary education, Grades, Free rider, Group assessments

Paper type Research paper



© Stephen Dix. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Asia Pacific Journal of Marketing
and Logistics
Vol. 36 No. 7, 2024
pp. 1767-1781
Emerald Publishing Limited
1355-5855
DOI 10.1108/APJML-09-2023-0858

Background

University students are co-creators of the outcomes that manifest from their personal educational journey. Each student is responsible to optimise their time and effort in order to maximise their learning and personal growth. In setting one or more assessments, instructors provide the mechanism to grade each student's progress in relation to a specified task.

Besides assuming responsibility for the quality of their assessment submissions, students may be exposed to other potential touchpoints that impact their own and others' grade outcomes. For example, through formative evaluation of peer submissions, students may advance their personal learning insights (Ocampo *et al.*, 2023) and so increase their own grades. Or, when invited to evaluate each member's effort within a group task, students may impact the grades of fellow members (Pond *et al.*, 2007).

In a study of preferences amongst various assessment methods (Neto *et al.*, 2022), students rated group work lowest in fairness and accuracy compared to other assessment options. The authors speculate that this belief forms when the same grade is awarded to all group members, despite sizable differences in individual effort levels.

This points to a key challenge. Tertiary educators overseeing group assessments face a dilemma. "Do I allocate the same overall grade to all members, or do I seek to moderate each individual member's grade to reflect personal effort?"

Literature review

In the absence of a directive from the educational institution, moderation of group work is typically left to the academic coordinator (Khuzwayo, 2018). From an academic instructor's perspective, allocation of the same grade to all members presents an attractive option for three reasons. Firstly, moderation of grades soaks up valuable time during a busy grading season. Secondly, the coordinator averts potential persistent complaints from low-rated, disgruntled students. Thirdly, in high-functioning groups where all members make equitable contributions, there is no need for a moderation process.

On the other hand, hard-working students may become frustrated when a group member, who makes little or no contribution, receives an equivalent grade. Students have a sense of fairness across their own and others' contribution to a group task (Sridharan *et al.*, 2018a) and would value the opportunity to have their say. The primary indicator of the success for the moderation of group projects is how the students themselves rate its value. Studies by Pond *et al.* (2007) and Crockett and Peter (2003) attest to the perceived validity of group moderation with students claiming the process to be fair. Abernethy and Lett (2005) report that the two most important drivers of *teamwork experience* are the grade received and the moderation of grades to reflect the individual contributions to the task. Moreover, capable students are generally welcoming of broadening moderation across group tasks and tend to express strong negative opinions about slack group members (Abernethy and Lett, 2005).

Slack contributors are known as a *social loafers* when their contribution is below expectation (Abernethy and Lett, 2005). Both weak and strong students working below their potential qualify as *social loafers* (Joyce, 1999). If they make no contribution to a group project, slack students are called *free riders* (Abernethy and Lett, 2005). It is difficult for instructors to identify free riders and social loafers as they may present themselves positively to tutors but do little or no work behind the scenes (Wilmot and Crawford, 2007).

A pressing concern for educators is that a shared group mark does not reflect the underperforming student's authentic contribution (Barfield, 2003; Cheng and Warren, 2000; Houldsworth and Mathews, 2000; Khuzwayo, 2018; Willmot and Crawford, 2007; Zhang and Ohland, 2009). Having *free riders* waltz away with a group mark that pushes their final grade into the "pass zone" represents a *service failure* within our education system. In the absence of

a moderation process, *social loafers* and *free riders* are effectively rewarded for academic misconduct.

It may be argued that free riding is the group assessment equivalent of what contract cheating constitutes within an individual assessment. In the absence of moderating group grades, educators inadvertently enable underperforming students to benefit from the work of others. Moderation of grades is the only means to eliminate this potential transgression. Once the effort level of each individual in the group is determined, all group members' grades may be moderated to reflect the result that they deserve. Moderating grades to mirror each student's contribution applies the "adjustment" necessary to deter *social loafers* and *free riders* (Abernethy and Lett, 2005).

Aim of this paper

The aim is to generate a peer-driven survey instrument to evaluate each group member's contribution to a group project. Where a member's contribution falls below the group average, that student's grade is adjusted down proportionally.

This purpose is already served by existing programmes such as Peer Assessment Pro, CATME and SparkPLUS. These are extensive programmes targeting groupwork effectiveness by means such as advancing group engagement, instating peer/instructor feedback and encouraging self-reflection (Catme, 2023; Peer Assessment Pro, 2023; Sparkplus, 2023). Whilst these initiatives bring clear benefits to the groupwork experience, they are complex, include sophisticated calculation and are potentially time-consuming for both students and instructors (Ohland *et al.*, 2012).

This paper focuses more narrowly on the issue of moderation, seeking to develop a streamlined, transparent and open-source survey instrument (TANDEM ©) to fairly moderate individual grades within a tertiary group project. Given TANDEM'S © dominant focus on grade moderation, it offers a hands-on solution for instructors whose key aim is to appropriately award students within group assessments. This is an expedient instrument that can be applied when and where needed at the discretion of the course coordinator. Whereas licence fees and formal university agreements typically associate with existing instruments (Catme, 2023; Peer Assessment Pro, 2023; Sparkplus, 2023), TANDEM © can be applied as needed by individual instructors. Minimal effort is required for students to complete 2 *paper and pen* surveys, each taking no more than 5 min. A modest time commitment is required for instructors to enter survey data into an excel spreadsheet (available from the author).

This paper draws on the literature to support the structure and operationalisation of the TANDEM © survey instrument. Importantly, TANDEM © is an open-source instrument that can be adapted and applied at no cost by any instructor within a tertiary group assessment context. With minimal administrative load on instructors and students, the instrument can be applied as a blanket moderation across the entire student cohort or as a means to fairly moderate individual grades in response to a request from a single group in conflict.

Foundation for an equitable peer evaluation process

If the underlying rationale for applying moderation to group projects is acceptable, the next step is to determine who is best placed to rate each member's contribution to the task. Input on each individual's contribution to the task may be gathered from a variety of sources including the unit coordinator, group leader, group members or assigned observers.

Based on the premise that students are best placed to rate one another's work (Millis and Cottell, 1998), the majority of studies gather data directly from peers (Brown, 1995, Mohammed and Angell, 2004; Brooks and Ammons, 2003; Lejk and Wyvill 2001a;

Goldfinch, 1994; Willmot and Pond, 2012; Zhang and Ohland 2009). Peer evaluation may be obtained from group members in the form of quantitative and/or qualitative data. For quantitative data collection, group members typically rate peer contribution via a survey (Mohammed and Angell, 2004; Brooks and Ammons, 2003; Lejk and Wyvill 2001a; Goldfinch, 1994). In qualitative settings, instructors typically set up a discussion opportunity for members to express their viewpoint regarding individual contributions to the final submission (Scager *et al.*, 2017).

Willmot and Pond (2012) criticise the involvement of student members to determine contribution scores. They centre their concerns on biases stemming from personality clashes, giving friends inflated ratings as well as thoughtless and vindictive ratings. Although these are valid sources of potential error, it is argued that they do not outweigh the underlying value of drawing input from those closest to the task. Peers are closest to the action. Outside of independent observers and recorded meetings (both impractical methods for ongoing use), peers are the only witnesses to how the group task unfolds. Thus, educators should proceed to apply peer evaluation of member contributions and, at the same time, seek to minimise the potential for bias.

It's notable that the mere inclusion of some form of peer evaluation into the moderation process may not necessarily lead to an accurate measure of individual contribution. A peer evaluation approach may be effective on one dimension but fall short in other ways. For example, a novel approach proposed by Abernethy and Lett (2005) gives peers the right to "fire" a group member. To alert the student to non-performance, a group member sends an initial email to the non-performing student and includes the unit coordinator into the communication. This email outlines the specifics of what the recipient is required to do and by when. If the non-performer fails to meet the requirements, a second email notifies the student that s(he) is fired, resulting in a zero grade. Whilst this method applies the harshest penalty to *free riders* (a zero grade), *social loafers* can still fly below the radar by meeting the group's basic expectations.

Criteria for building an effective moderation instrument

This paper strives to develop a streamlined student-focused survey to fairly moderate group grades for members whose overall peer rating falls below the group average threshold. Several core tenets drawn from the literature are embedded into the development and structure of the TANDEM © peer evaluation instrument.

Early implementation and clear communication. Early implementation and clear communication (Goldfinch, 1994) of the moderation approach is essential to its success. Brooks and Ammons (2003) report that early evaluation and specific feedback led to a reduced free-rider problem.

In order to address this criterion, early and clear communication is imperative to the success of the peer evaluation process. Not only should the process and application be detailed in the Unit Information, but students should be briefed during the first class. In order to reinforce the process, instructors should offer a refresher briefing at the start of the group project as well as reminders prior to survey occasions.

Multiple evaluation points. More than one evaluation of peer contribution is critical in addressing the problem of social loafing or free riding, thus steering the group towards a fair outcome (Abernethy and Lett, 2005; Aggarwal and O'Brien, 2008). A single, summative (end-of-task) application of a peer evaluation instrument may serve to moderate grades of underperformers (Sridharan *et al.*, 2018b). However, over two consecutive peer evaluations, a significant drop in the variance across evaluation scores can evidence a reduction the free-rider problem (Brooks and Ammons, 2003).

Ideally, students complete the peer evaluation survey mid-task (formative) and then again at the end of the group task (summative). It is proposed herein that a mid-task, phase 1 survey (formative) provides the instructor an opportunity to inform students on how their peers rate their contribution to date. Importantly, students are reminded that mid-task (formative) ratings have zero impact on their final grades. Only the end-of-task, phase 2 (summative) ratings are used to moderate group members' final grades, as necessary.

Confidential completion of peer evaluation surveys. Peer ratings completed in private show greater diversity compared to when completed openly in the presence of the group (Lejk and Wyvill, 2001b). The presumption is that, in private, students are not subject to direct "social obligation" when rating their peers. Moreover, Li (2017) reports that anonymity reduces peer pressure and increases student comfort levels within peer evaluation projects. Finally, Sidharan *et al.* (2018a) report that students who value the anonymity of peer evaluation also recognise a reduction in the incidence of the "free-riding problem".

An effective instrument requires that confidentiality is guaranteed. In face-to-face settings, students are expected to complete the TANDEM © survey confidentially using several remote table locations across the classroom. Once completed, students fold the form and personally hand it to the instructor. In online settings, individuals are emailed to complete the survey and reply directly to the instructor by a specified date.

Customise the instrument to a specific learning target. The peer evaluation criteria should ideally steer towards specific learning outcomes that apply to the group task at hand (Zhang and Ohland 2009). Within the context of students' contributions to a group project, survey dimensions may be aligned with a specified discipline area in mind. For example, contribution-to-task by nursing students may carry different expectations from those of business students engaged in group work. Discipline instructors should gather to discuss and determine key drivers/dimensions of student contribution to the group task. Ideally, these dimensions should be applied consistently to peer evaluation survey instruments across all courses within the same discipline.

Inclusion of category versus holistic peer evaluation questions. Category based peer evaluation occurs where group members are rated across multiple underlying dimensions (Kilic and Cakan 2006). Metrics such as team skills, time management, communication and technical skills are typical dimensions applied to peer evaluation (Kilic and Cakan 2006; Ohland *et al.*, 2012; Zhang and Ohland 2009).

In this paper, by way of providing an example of a faculty-based approach, four generic dimensions are shown in the TANDEM © survey (Appendix 1). These were harvested through a collaborative "think-tank" amongst academics within the business faculty. The academic group involved favoured strong behavioural dimensions to rate students' contribution to the project. These dimensions are not only visible to peers, but also associate with behaviours that are valued within a business context. Whilst these dimensions were applied to a business context, it is recognised that the ideal mix of dimensions may vary across diverse academic disciplines.

The approach applied in the TANDEM © instrument requires students to rate self and peers on a scale of 1–10 for each of the following four equal-weighted category dimensions:

- (1) Attended group meetings consistently (meetings can be actual and/or virtual)
- (2) Contribution to the project
- (3) Is a team player
- (4) Completed work on time

Holistic peer evaluation occurs when group members rate overall contribution via a single measure. An example of this is asking respondents to allocate a certain number of points

amongst group members to represent each individual's contribution to the task (Ohland *et al.*, 2012). A question based on this approach is included in the TANDEM © instrument in the following form:

- (1) How might you pay them? You have a total of \$100 to split amongst your members to show each person's overall contribution to the project. How much would you pay each member?

This question has been adapted by allocating dollars rather than points. This brings a practical slant in that monetary reward suggests a context of how much each individual would "earn" if their contribution was paid for. Moreover, since students are required to exclude themselves from the dollar allocation response, the potential for personal bias is lessened (Goldfinch, 1994). Ohland *et al.* (2012) suggest that since a holistic measure does not promote teamwork nor enable instructor feedback, it brings little value to a broad teamwork system. However, in the context of this streamlined, moderation-only focus, a holistic measure may prove to be a worthy addition to the TANDEM © survey. With both category dimensions ratings as well as a holistic measure in place, TANDEM © captures two independent peer-evaluation measures. Preliminary observations from the inclusion of this holistic measure will be outlined in the latter part of this paper.

Lejk and Wyvill (2001a) report that peer evaluations based on category dimensions may be better suited to formative feedback, whilst holistic peer evaluation better aligns with summative feedback. The purpose of the mid-task phase 1 peer evaluation survey (Appendix 1) is to provide formative feedback to students. Students are awarded an average score for each category dimension based on their group members' ratings. In providing this feedback to students, this constitutes a call-to-action for underperforming students to lift their contribution to the task.

Students rating group members should also rate themselves. The literature is mixed on this issue. Some studies report that students should rate themselves in addition to rating their group members (Goldfinch, 1994). However, when generous students give higher ratings to others than they receive from their group, their total score is downsized relative to their peers' total scores. Goldfinch (1994) argues that by including self-rating into the mix, these students may inflate their own total score to offset the potential problem of relative downsizing.

Other researchers argue that low-performing students may unjustifiably inflate their own score ("strategic self-rating") when self-rating is included (Willmot and Crawford, 2007; Lejk and Wyvill, 2001a; Zhang *et al.*, 2008). In a study assessing the reliability of self and peer ratings, Zhang *et al.* (2008) warn that "students can rate themselves very differently from the way they rate others".

In mid-task phase I (formative) of the TANDEM © Peer Evaluation Instrument (Appendix 1), self-rating is required across the four Category Peer Evaluation questions. However, this self-rating score is not included into the analysis of peer contribution. In this approach, the self-rating serves only as a marker to stimulate student reflection. A student is called to consider his or her personal contribution score to the task relative to the scores he or she awarded to peers. However, to avoid strategic self-rating, this score is omitted from the ensuing analysis.

The end-task phase 2 (summative) survey (Appendix 2) duplicates phase 1 but includes one additional holistic question. This holistic question requires that students split \$100 amongst group members to fairly compensate for their overall contribution to the project. This presents a single metric that may serve as an alternative summative measure of peer contribution to the task. To eliminate potential bias arising from "strategic self-rating", the survey respondent is directed not to include him or herself in this split.

Methods, analysis and discussion

Development of a peer evaluation instrument for group projects

The TANDEM © Peer Evaluation instruments seek to determine student contribution and, where necessary, apply a fair grade modification ([Appendixes 1 and 2](#)). At the same time, the intent is to minimise the additional workload placed on the instructor.

Phase I (mid-task): interim indication of student contribution. As previously mentioned, the phase 1 survey ([Appendix 1](#)) requires each student to rate all group members on the category dimensions, including him or herself. Recall that the responding member's self-evaluation is not included into the outcome analysis.

Phase 1 survey is confidentially completed by all group members at the mid-way stage of the project. Recall that each group member is rated by all other members against four dimensions ([Appendix 1](#)): Meeting Attendance; Contribution to date; Team Player: Completes Work on Time. The rating is based on a scale of 1–10 (1 or 2 = Well below Expectation; 3 or 4 = Below Expectation; 5 or 6 = Meets Expectation; 7 or 8 = Above Expectation; 9 or 10 = Well above Expectation).

Students may receive feedback as a formative cue for their contribution to the group task to date. In order to minimise administration time, instructors may opt to email feedback to underperforming students only. This feedback may comprise three key performance indicators across the 4 dimensions:

- (1) Your self-rating.
- (2) Your average score as rated by your group members.
- (3) The average score obtained by all other group members.

An example of formative feedback from the Phase 1 survey emailed to an underperforming student is shown in [Table 1](#). This notification is a “call-to-action” to encourage under-active or non-active students to lift their contribution to the task.

Phase 2 (end-of-task): final evaluation of student contribution. During phase 2, students are given a second (and final) opportunity to rate each group member's contribution based on four category dimensions as well as on a single holistic measure. This is completed at the end of the project task. Students are reminded that the outcome of the analysis will potentially

	Meeting attendance	Team player	Completes work on time	Contribution to the project to date	Average rating
Your self-rating	7/10	8/10	7/10	8/10	7.5/10
Your average rating by your group members	5/10	6/10	4/10	3/10	4.5/10
Average rating obtained by your group members	7/10	9/10	8/10	8/10	8/10

Note(s): Your group members have rated you for the group project to date. Your group members have rated you across the four performance areas with an average of 4.5/10. This falls below the average rating for all your group members (8/10). [Note that these do not include self-ratings]

Take note of the areas for which your group members rate you lowest. Aim to improve these areas to make a stronger contribution to the project going forward

Your personal average rating is 3.5 points below the group average ($8 - 4.5 = 3.5$). This is equivalent to contributing 44% below your group's average. If your group members see no improvement going forward, they may rate you at this same level in the Phase 2 survey (end of the project). In this case, you will stand to lose 44% of the grade that is awarded to your group for this project

Source(s): Created by author

Table 1.
Example of Phase 1
(mid-project) feedback
to underperforming
student

affect the grades of group members. The Phase 2 survey instrument replicates the Phase 1 survey process. Once again, students are required to rate themselves and their group members across the same four category dimensions.

Students who perform at or above the group's overall average rating receive the full extent of the grade awarded to the group. The TANDEM © moderation only applies to students whose personal overall average score falls below the group's average rating (self-ratings are excluded from the analysis). The underperforming student is proportionally penalised by the percentage that his or her overall rating falls below the group average (Table 2).

Given the final grade of 80% awarded to the group, those students at or above the group average rating (94) are awarded the full grade (80%). Those below the 94 threshold are moderated down in proportion to the deficit of their personal overall average rating relative to the group overall average rating. For example, based on the scores presented by his three group members, Sam's average rating is 87. The overall average rating across the group is 94. Since Sam's score falls below 94, his grade is moderated down proportionally. Sam is awarded $87/94 = 0.925$ of the project grade. Given the group project grade is 80%, Sam receives a personal grade of 74% ($0.925 \times 80 = 74$).

Additional holistic measure in phase 2 survey

As previously mentioned, there is one additional holistic question posed in the Phase 2 survey instrument. This may augment or replace the four category dimensions with a single summative measure of peer contribution. Recall that it has been noted in the literature that a single holistic question may be better aligned to peer feedback within a summative context (Lejk and Wyvill, 2001a). If this measure proves to be better suited to summative settings, instructors may further reduce administrative load by replacing aggregate ratings across four categories with a single holistic measure.

Recall that this holistic measure requires each member to divide \$100 amongst peers to express their contribution to the task. Recall too that the respondent does not include him or herself into the calculation to avoid potential biased ratings by underperforming or strategic group members.

It is recommended that, at this point in time, the category dimensions approach is used to moderate group member's grades in Phase 2. Moreover, in applying the same four measures to both phases 1 and 2 of the TANDEM © survey, this provides greater uniformity across the process.

Group member	Average rating across 4 dimensions	Moderation scaling method	Group project grade	Post-moderated final grade
Person 1	97	<i>Students with an overall average rating at or above 94: No moderation occurs</i>	80%	80%
Person 2	93			$80 \times 93/94 = 79\%$
Person 3	98			80%
Person 4 (Sam)	87			$80 \times 87/94 = 74\%$
<i>Group average rating</i>	94	<i>Students with an overall average rating below 94: The grade is moderated down in proportion to the % below the group mean</i>		

Table 2. Example of Phase 2 (end-of-project) moderation of grades
Source(s): Created by author

To date, monitoring of the correlation between the holistic metric and category dimensions suggests a very strong positive correlation. Across a sample of student ratings applying this exact method ($n = 33$), a correlation coefficient of $r = 0.89$ occurred between the holistic rating and average category dimensional ratings.

Table 3 outlines an example of how the holistic (dollar allocation) measure may be applied to moderating final grades.

The single holistic measure typically applies a greater penalty to underperforming students. Recall that Sam's phase 2 grade was moderated to 74% based on the four category dimensions. However, it moderates down to 67% based on the single holistic measure. Within a four-person group where one member rates three others, an equal split suggests that each member receives an *unbiased mean* of \$33.33. Therefore, Sam's average allocation (\$27.78) is \$5.55 below the mean. This equates to a dollar allocation of 16.65% below the unbiased mean of \$33.33. Accordingly, his final grade is proportionally scaled down by 16.65%. If the group earns a grade of 80% for the group project, Sam's moderated grade reduces to 67% ($80\% \times 27.78/33.33 = 67\%$).

Further investigation is required before instructors opt to exclusively apply the holistic grade moderation across phase 2 (summative) of the peer evaluation.

Concluding comments

There is no one perfect method of determining contribution to a group task (Zhang *et al.*, 2008), but it has been noted that the mere act of including a moderation component into the course drives greater accountability and student investment into the group task (Brown, 1995). A student who risks a lower grade driven by poor peer ratings is incentivised to invest more into the task. Clearly, with a meaningful moderation instrument in place, those privy to *who did what* are empowered to re-calibrate the grade(s) of persistent social loafers and free riders.

Implications for practitioners

TANDEM © serves predominantly as a grade moderation instrument for underperforming students based on peer evaluations of group member contribution. Notably, there is a secondary benefit in the form of an alert to those students at risk of a downgrade. This serves as a mid-project call-to-action to underperforming students noting the potential of a grade reduction at current effort levels (Phase 1). Finally, a grade moderation is actually applied at the end of the project to any underperformers (Phase 2).

The key advantage of TANDEM© lies in its simplicity and transparency. A variety of moderation methods are built into contemporary alternative instruments that propose complex algorithms (Ko, 2014) that risk presenting as “black box” solutions from a student perspective. Although these offerings include several group performance benefits, they place considerable workload demands on students and instructors. Moreover, these are typically commercially driven products that target university decision makers.

	Person 1	Person 2	Person 3	Sam	Average \$
Person 1		\$33.33	\$50	\$30	\$37.78
Person 2	\$35		\$30	\$30	\$31.66
Person 3	\$35	\$33.33		\$40	\$36.11
Person 4 (Sam)	\$30	\$33.33	\$20		\$27.78
					\$33.33 (<i>unbiased mean</i>)

Source(s): Created by author

Table 3.
Example of dollar
allocations amongst 4
group members
(holistic measure)

In providing a streamlined and confidential process for peer-rated moderation, TANDEM© draws extensively from the literature. Its intent is to enable an open-source moderation instrument for individual instructors to activate as required. Within the context of a modest administrative load, TANDEM © focuses primarily on providing a moderation outcome, making this instrument an expedient application alternative for instructors.

When applied across a major or a discipline, students will benefit greatly from having the same moderation instrument consistently applied to all component courses. Not only will they appreciate the opportunity to rate their peers' contribution, but high performing students in particular will value that grades are consonant with the effort applied by each individual involved in the group project.

TANDEM © may be applied to all groups across the course or can selectively be applied to groups that report moderate, strong or extreme levels of conflict ([Appendixes 1 and 2](#)).

The analysis of a group of four students is drawn from an Excel spreadsheet and is shown in [Appendix 3](#).

Limitations and further research opportunities

Ideally, TANDEM© should be applied to group sizes across four or five members. However, its applicability to three-person groups is yet to be established. The instrument has been trialled only amongst under-graduate students. It is yet to be applied across post-graduate groups and within online learning environments. There is also opportunity to apply the instrument into different countries to compare uptake and ratings among culturally-diverse student cohorts.

The additional holistic “dollar allocation” question is added to the phase 2 summative survey as a single-item verification of the moderation outcome. However, this may give rise to further research to determine whether this metric serves as a reliable surrogate for the four-item alternative.

There is no apparent investigation into how students themselves respond to a moderated grade. Motivated students are clearly in favour of peer moderated grades ([Abernethy and Lett, 2005](#); [Crockett and Peter, 2003](#); [Pond et al., 2007](#)). However, there is a dearth of studies that capture how students rate the fairness of moderated grades amongst group members. Although there are obvious constraints around confidentiality, it would be enlightening to determine student responses to peer-adjusted grade allocations amongst their group based on the TANDEM © instrument.

Finally, as mentioned, the mere presence of a peer moderation instrument may serve to encourage *social loafers* and *free riders* to raise their effort levels. Further research into this area of interest may offer insight into how motivated versus demotivated students may respond to the presence versus absence of a peer evaluation instrument.

References

- Abernethy, A. and Lett, W. (2005), “You are fired! A method to control and sanction free riding in group assignments”, *Marketing Education Review*, Vol. 15 No. 1, pp. 47-54, doi: [10.1080/10528008.2005.11488891](#).
- Aggarwal, P. and O'Brien, C.L. (2008), “Social loafing on group projects structural antecedents and effect on student satisfaction”, *Journal of Marketing Education*, Vol. 30 No. 3, pp. 255-264, doi: [10.1177/0273475308322283](#).
- Barfield, R.L. (2003), “Students’ perceptions of satisfaction with group grades and the group experience in the college classroom”, *Assessment and Evaluation in Higher Education*, Vol. 28 No. 4, pp. 355-370, doi: [10.1080/0260293032000066191](#).

- Brooks, C. and Ammons, J. (2003), "Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments", *Journal of Education for Business*, Vol. 78 No. 5, pp. 268-272, doi: [10.1080/08832320309598613](https://doi.org/10.1080/08832320309598613).
- Brown, R.W. (1995), "Autorating: getting individual marks from team marks and enhancing teamwork", *Proceedings from 1995 ASEE/IEEE Frontiers in Education Conference*, Vol. 2, p. 3c2-15.
- Catme (2023), Catme website, available at: <https://www.info.catme.org/> (accessed 16 November 2023).
- Cheng, W. and Warren, M. (2000), "Making a difference: using peers to assess individual students' contributions to a group project", *Teaching in Higher Education*, Vol. 5 No. 2, pp. 243-255.
- Crockett, G. and Peter, V. (2003), "Peer assessment in a second year macroeconomics unit", in *The Higher Education Academy*, Economics Network Extended Case Study.
- Goldfinch, J. (1994), "Further developments in peer assessments of group projects", *Assessment and Evaluation in Higher Education*, Vol. 19 No. 1, pp. 29-35, doi: [10.1080/0260293940190103](https://doi.org/10.1080/0260293940190103).
- Houldsworth, C. and Mathews, B. (2000), "Group composition, performance and educational attainment", *Education and Training*, Vol. 42 No. 1, pp. 40-53, doi: [10.1108/00400910010317086](https://doi.org/10.1108/00400910010317086).
- Joyce, W. (1999), "On the free-rider problem in cooperative learning", *Journal of Education for Business*, Vol. 74 No. 5, pp. 271-274, doi: [10.1080/08832329909601696](https://doi.org/10.1080/08832329909601696).
- Khuzwayo, M. (2018), "Assessment of group work in initial teacher education and training", *South African Journal of Education*, Vol. 38 No. 2, pp. 1-11.
- Kilic, G.B. and Cakan, M. (2006), "The analysis of the impact of individual weighting factor on individual scores", *Assessment and Evaluation in Higher Education*, Vol. 31 No. 6, pp. 639-654, doi: [10.1080/02602930600760843](https://doi.org/10.1080/02602930600760843).
- Ko, S.S. (2014), "Peer assessment in group projects accounting for assessor reliability by an iterative method", *Teaching in Higher Education*, Vol. 19 No. 3, pp. 301-314, doi: [10.1080/13562517.2013.860110](https://doi.org/10.1080/13562517.2013.860110).
- Lejk, M. and Wyvill, M. (2001a), "Peer assessment of contributions to a group project: a comparison of holistic and category-based approaches", *Assessment and Evaluation in Higher Education*, Vol. 26 No. 1, pp. 61-72, doi: [10.1080/02602930020022291](https://doi.org/10.1080/02602930020022291).
- Lejk, M. and Wyvill, M. (2001b), "The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: a quantitative study of secret and agreed assessments", *Assessment and Evaluation in Higher Education*, Vol. 26 No. 6, pp. 551-561, doi: [10.1080/02602930120093887](https://doi.org/10.1080/02602930120093887).
- Li, L. (2017), "The role of anonymity in peer assessment", *Assessment and Evaluation in Higher Education*, Vol. 42 No. 4, pp. 645-656, doi: [10.1080/02602938.2016.1174766](https://doi.org/10.1080/02602938.2016.1174766).
- Millis, B.J. and Cottell, P.G. Jr (1998), *Cooperative Learning for Higher Education Faculty*, American Council on Education/Oryx Press, Phoenix, AZ.
- Mohammed, S. and Angell, L. (2004), "Surface and deep level diversity in workgroups: examining the moderating effects of team orientation and team process on relationship conflict", *Journal of Organizational Behaviour*, Vol. 25 No. 8, pp. 1015-1039, doi: [10.1002/job.293](https://doi.org/10.1002/job.293).
- Neto, J., Neto, F. and Furnham, A. (2022), "Predictors of students' preferences for assessment methods", *Assessment and Evaluation in Higher Education*, Vol. 48 No. 4, pp. 556-565, doi: [10.1080/02602938.2022.2087860](https://doi.org/10.1080/02602938.2022.2087860).
- Ocampo, J.C., Panadero, E. and Diez, F. (2023), "Are men and women really different? The effects of gender and training on peer scoring and perceptions of peer assessment", *Assessment and Evaluation in Higher Education*, Vol. 48 No. 6, pp. 760-776, doi: [10.1080/02602938.2022.2130167](https://doi.org/10.1080/02602938.2022.2130167).
- Ohland, M.W., Loughry, M.L., Woehr, D.J., Bullard, L.G., Felder, R.M., Finelli, C.J., Layton, R.A., Pomeranz, H.R. and Schmucker, D. (2012), "The comprehensive assessment of team member effectiveness: development of a behaviourally anchored rating scale for self- and peer evaluation", *Academy of Management Learning and Education*, Vol. 11 No. 4, pp. 609-630, doi: [10.5465/amle.2010.0177](https://doi.org/10.5465/amle.2010.0177).

- Peer Assessment Pro (2023), "Peer assessment pro website", available at: <https://peerassesspro.com/> (accessed 16 November 2023).
- Pond, K., Coates, D. and Palermo, O. (2007), "Student experiences of peer review marking of team projects", *International Journal of Management Education*, Vol. 6 No. 2, pp. 30-43, doi: [10.3794/ijme.62.190](https://doi.org/10.3794/ijme.62.190).
- Scager, K., Boonstra, J., Peeters, T., Vulperhorst, J. and Wiegant, F. (2017), "Collaborative learning in higher education: evoking positive interdependence", *Life Science Education*, Vol. 15 No. 4, p. ar69, doi: [10.1187/cbe.16-07-0219](https://doi.org/10.1187/cbe.16-07-0219).
- Sparkplus (2023), Sparkplus website, available at: <https://www.sparkplus.com.au/> (accessed 16 November 2023).
- Sridharan, B., Muttakin, M. and Mihret, D. (2018a), "Student's perceptions of peer assessment effectiveness: an explorative study", *Accounting and Education*, Vol. 27 No. 3, pp. 1-27, doi: [10.1080/09639284.2018.1476894](https://doi.org/10.1080/09639284.2018.1476894).
- Sridharan, B., Tai, J. and Boud, D. (2018b), "Does the use of summative peer assessment in collaborative group work inhibit good judgement?", *Higher Education*, Vol. 77 No. 5, pp. 853-870, doi: [10.1007/s10734-018-0305-7](https://doi.org/10.1007/s10734-018-0305-7).
- Willmot, P. and Crawford, A. (2007), "Peer review of team marks using a web-based tool: an evaluation", *Engineering Education: The Journal of Higher Education Academy Engineering Subject Centre*, Vol. 2 No. 1, pp. 59-66, doi: [10.11120/ened.2007.02010059](https://doi.org/10.11120/ened.2007.02010059).
- Willmot, P. and Pond, K. (2012), "Multi-disciplinary peer-mark moderation of group work", *International Journal of Higher Education*, Vol. 1 No. 1, pp. 2-13, doi: [10.5430/ijhe.v1n1p2](https://doi.org/10.5430/ijhe.v1n1p2).
- Zhang, B. and Ohland, W.M. (2009), "How to assign individualized scores on a group project: an empirical evaluation", *Applied Measurement in Education*, Vol. 22 No. 3, pp. 290-308, doi: [10.1080/08957340902984075](https://doi.org/10.1080/08957340902984075).
- Zhang, B., Johnston, L. and Kilic, G.B. (2008), "Assessing the reliability of self and peer rating in student group work", *Assessment and Evaluation in Higher Education*, Vol. 33 No. 3, pp. 329-340, doi: [10.1080/02602930701293181](https://doi.org/10.1080/02602930701293181).

Further reading

- Cheng, W. and Warren, M. (2000), "Making a difference: using peers to assess individual students' contributions to a group project", *Teaching in Higher Education*, Vol. 5 No. 2, pp. 243-255, doi: [10.1080/135625100114885](https://doi.org/10.1080/135625100114885).

Appendix 1

TANDEM peer evaluation (Phase 1): peer evaluation of contribution to group project (mid-task)

- (1) This is a confidential survey. Your actual ratings are not disclosed to anyone other than your Instructor(s).
- (2) Rate each group member's overall contribution to the project so far.
- (3) Please provide a fair rating. Do not let friendship connections or personality clashes influence your scores.
- (4) In Phase 1, your rating *will not affect your grades and will not affect any of your group's grades*.
- (5) You will only be notified if your personal rating falls below your group's average rating.

How did your group members perform?

Guide to rating your group members:					
Here's how to rate your members. For example, a member who meets expectation is allocated 5 or 6 points (out of 10)					
Rate out of 10 points	Well below expectation	Below expectation	Meets expectation	Above expectation	Well above expectation
	1 or 2	3 or 4	5 or 6	7 or 8	9 or 10

Rate each group member out of 10 points in ALL columns (1 to 4)

Write the full name of each group member below:	1. Attended group meetings consistently	2. Contribution to the project to date	3. Is a team player	4. Completed work on time
1.				
2.				
3.				
Rate yourself here				

One the scale below (1 = no conflict; 10 = extreme conflict), circle the number that best describes the level of conflict within your group:

No conflict		Mild conflict		Moderate conflict		Strong conflict		Extreme conflict	
1	2	3	4	5	6	7	8	9	10

Any comments?

Source(s): Created by author

Appendix 2

TANDEM peer evaluation (Phase 2): peer evaluation of contribution to group project (end-of-task)

- (1) This is a confidential survey. Your actual ratings are not disclosed to anyone other than your Instructor(s).
- (2) Rate each group member's overall contribution to the entire project.
- (3) Please provide a fair rating. Do not let friendship connections or personality clashes influence your scores.
- (4) Your rating *will impact your group members' grades*.

(5) Any individual with an average rating below the average group rating receives a proportional grade reduction.

How did your group members perform?

Guide to rating your group members:

Here's how to rate your members. For example, a member who meets expectation is allocated 5 or 6 points (out of 10)

Rate out of 10 points	Well below expectation	Below expectation	Meets expectation	Above expectation	Well above expectation
	1 or 2	3 or 4	5 or 6	7 or 8	9 or 10

Rate each group member out of 10 points in columns 1 to 4

Write the full name of each group member below:	1. Attended group meetings consistently	2. Contribution to the project to date	3. Is a team player	4. Completed work on time	How might you pay them? You have a total of \$100 to split fairly among your members for each person's overall contribution to the project. How much would you pay each member? DO NOT INCLUDE YOURSELF
1.					
2.					
3.					
Rate yourself here					Do not pay yourself (leave this blank)

One the scale below (1 = no conflict; 10 = extreme conflict), circle the number that best describes the level of conflict within your group:

No conflict		Mild conflict		Moderate conflict		Strong conflict		Extreme conflict	
1	2	3	4	5	6	7	8	9	10

Any comments?

Source(s): Created by author

Appendix 3

University
students'
grades for
group projects

1781

	RATINGS BY TEAM MEMBERS					GROUP GRADE	INDIV GRADE (RATINGS)	INDIV GRADE (DOLLARS)
	meeting	contribution	team player	complete work	payment			
3	Person 1	7	7	9	4	80%	80%	80
4	Person 2	10	9	9	8	80%	79%	76
5	Person 3	10	10	9	8	80%	80%	80
6	Person 4	6	7	9	8	80%	74%	66.7
7		10	10	9	8			
8								
9	Person 1	8	9	10	9	33.33		
10	Person 2	9	9	10	9			
11	Person 3	9	9	10	9	33.33		
12	Person 4	8	9	10	9	33.33		
13		9	9	10	9			
14								
15	Person 1	6	8	8	7	50		
16	Person 2	6	7	8	7	30		
17	Person 3	8	10	10	10			
18	Person 4	5	5	6	7	20		
19		6	8	8	7			
20								
21	Person 1	6	7	5	7	30		
22	Person 2	6	5	5	6	30		
23	Person 3	6	9	4	7	40		
24	Person 4	6	6	6	7			
25		6	9	5	7			

Source(s): Created by author

Figure A1.
Excel spreadsheet to
determine moderated
outcomes by category
dimension (ratings)
and holistic measure
(dollars)

Corresponding author

Stephen Dix can be contacted at: steve.dix@cbs.curtin.edu.au

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com