Students' Test Satisfaction Goals are Well Calibrated to Their Test Scores

Abigail R. Vild[1], Maggie E. Wilson[1], & Christopher A. Was[1]
[1]Kent State University

Corresponding author: Christopher A. Was
Email: cwas@kent.edu

Theories of self-regulated learning suggest a positive link between knowledge monitoring accuracy (the ability to predict test performance) and performance on tests. Put differently, students who accurately monitor their knowledge of course content more efficiently regulate study of course materials. However, a plethora of literature indicates that most undergraduate students are overconfident when asked to predict performance on tests. Previous research has relied students making predictions about the score they expect to receive on an upcoming test. Although, intuitively asking students to predict their test scores as a measure of monitoring accuracy makes sense, students tend to be overconfident. This lack of accuracy is more pronounced in poorly performing students, as they tend to predict significantly higher tests scores than they receive. Several studies have attempted to increase student monitoring accuracy in hopes of increasing self-regulated learning and performance. Unfortunately, previous attempts have been unsuccessful. In the current study, we asked students to state their pride and satisfaction goals just before taking tests. Undergraduates were administered 11 tests during the semester. Before each test, students predicted their score, provided a score with which they would be satisfied (satisfaction goal), and with which they would be proud (pride goal). Students' predicted scores and pride goals were both significantly higher than actual test scores. However, satisfaction goals were lower than both predictions and pride goals and better reflected students' test scores. The results suggest that satisfaction goals accurately reflect the knowledge students have regarding the content to be covered on the test.

Keywords: metacognition, goal setting, self-regulation, possible selves

Metacognitive calibration is a critical quality of metacognitive monitoring during self-regulated learning (Pieschl, 2009). Calibration is defined as the accuracy of learners' perception of their own performance and is often measured as the difference between students' predicted performance (e.g., predictions of test scores) and actual performance (e.g., actual test scores). Unfortunately, students are often overconfident when predicting test performance and calibrations is especially poor for the poorest performers. Finding ways to improve students' perception of their future test performance, and in turn improve their self-regulated learning, is key to improving student success. The present study investigated whether undergraduate students' satisfaction goals and pride goals (Isaacson & Fujita, 2006) for course tests would be better calibrated to their test performance than their predictions of test performance. If goals are better calibrated to test scores than perhaps having students reflect upon their goals for a test would better help them to self-regulate their learning than having them attempt to predict their test scores.

The ability to accurately assess one's own knowledge (knowledge monitoring) is an important metacognitive process for students as accurate knowledge monitoring can help students improve study habits and self-regulate their learning, thereby increasing academic performance. Indeed, it is well-established in the metacognitive and educational literature that there is a striking difference between high performing and low performing students regarding the accuracy of test score predictions. Classroom research in university settings has repeatedly demonstrated this difference (e.g., Bol et al, 2005; Hacker et al, 2000; Isaacson & Was, 2010; Miller & Geraci, 2011a, 2011b; Nietfield et al, 2006). In the extant literature regarding metacognition, student participants are often asked to predict their scores on upcoming tests as a measure of their knowledge monitoring ability. Two common findings are that students' ability

64

to accurately predict test performance is related to actual performance on tests and calibration (the accuracy of their test score predictions) is often poor with a tendency for students to be overconfident. Typical findings also include a tendency for poorer performing students to be less well calibrated and more overconfident than higher performing students. Research in university classrooms has consistently demonstrated these findings (Bol et al, 2005; Hacker et al, 2000; Isaacson & Was, 2010; Leman et al, 2023; Miller & Geraci, 2011; Nederhand et al, 2021; Nietfield et al, 2006; Roeser et al, 1998).

There are common explanations for these findings. One explanation is that better performing students are better calibrated due to a measurement artifact within the data. Put differently, low performers have more room to make predictions well above their actual performance, whereas the highest performers do not have the room to make predictions much above their performance (Krueger & Mueller, 2002). A second interpretation is that the unskilled are unaware (Ehrlinger et al, 2008; Kruger & Dunning, 1999). This double-curse occurs because not only do poor performers lack the knowledge to correctly answer test questions, but they also lack enough knowledge to be aware they are not producing accurate responses (i.e., they are unable to judge the quality and accuracy of their responses). However, Miller and Geraci (2011) demonstrated that the lowest performers might not be as "blissfully incompetent" (Williams, 2004) as previously suggested. Miller and Geraci (2011) required undergraduate students to predict their letter grade on the first test of the semester immediately before the test was administered. Students were also asked to rate their confidence that their predicted score would be accurate (subjective confidence). This confidence rating of performance prediction is often referred to as a second order judgment (SOJ) (Dunlosky et al, 2005). Based on the unskilled are unaware hypothesis it would be expected that the poorest performing students (i.e., the unskilled)

65

would not be able to judge the accuracy of their test score predictions (i.e., they are unaware). As expected, Miller and Geraci (2011) found that when predicting test scores, the lowest performing students were significantly more overconfident than the highest performing students. But contrary to the unskilled are unaware hypothesis, lower performing students' SOJs were also lower than those of the best performing students. Put differently, the poorest performing students were aware that their test score predictions were not likely accurate.

Another explanation for students' overconfidence in their test score predictions, and even more so their pride goals, is the wishful thinking hypothesis (Stipek et al, 1984). Recall that the wishful thinking hypothesis proposed that students may fail to differentiate their outcome expectations from their desires. Thus, in the current study, test score predictions and pride goals may reflect students' hopes for their exam performance rather than their expectations for exam performance.

If poor performing students are aware that their test score predictions are inaccurate, what then are they using to make their test score predictions? Although there are numerous proposed explanations for the inaccuracy of their poor calibration (e.g., ego protection mechanisms, motivational explanations), there is a lack of understanding regarding the accuracy of their SOJs. The current study used a goal setting approach to determine if students have a better sense of how they will perform on a test than is reflected in their test score predictions. Our research question being, "Are students' performance (pride and satisfaction) goals better calibrated to actual test scores than the students' s predictions of their test scores?"

Ours is the first study that we are aware of to investigate whether students' satisfaction goals and pride goals are well calibrated to their actual exam scores. As part of a larger study, we asked students to predict their test scores, to report what test score they would be proud of, and

66

what test score would they need to achieve to be satisfied with their score. Undergraduates in an educational psychology course were administered 11 tests during a semester. Before each test, students predicted their score, provided a score with which they would be satisfied (satisfaction goal), and with which they would be proud (pride goal). We examined the data to determine if goals would be more accurate *predictors* of tests scores than actual predictions. We hypothesized that pride and satisfaction goals would both be better calibrated to test scores than test score predictions. We also hypothesized that although both goals would be better calibrated to test scores, pride goals would be less well calibrated (i.e., *overconfident*) than satisfaction goals.

Method

Participants

Seventy-five undergraduates enrolled in an educational psychology course participated in this study in exchange for course credit. Two sections of the course were taught by the same instructor during the same semester. Sixty-one percent of participants were female, and all were of sophomore or junior standing. Eight participants data were dropped from the analysis as their mean scores across tests were more than three standard deviations below the mean.

Design and Procedure

Participants were a part of a larger study investigating the relationships between self-regulated learning and metacognition. As part of the larger study, participants completed several questionnaires and measures of motivation (e.g., the Motivated Strategies for Learning Questionnaire, Garcia & Pintrich, 1995; the Learning and Study Strategies Inventory Weinstein et al, 2002) and the beginning of the semester which our study took place. Our interests in the current study only include students' ability to accurately predict test scores on the day of a test, and thus we do not report other data here.

During the semester in which students participated in the current study, students attended class on Monday, Wednesday, and Friday each week across 15 weeks. Mondays and Wednesdays were reserved for lecture whereas, beginning the second week, most Fridays served as the test day for each week. Test days involved completion of a test that tested knowledge of lecture material for that week. All tests were 35 multiple-choice items and were scored using automated scanner testing sheets. The Wednesday before the first test, the instructor provided students with a description of pride and satisfaction goals and asked them to think about their goals for each of the tests throughout the semester. Pride goals were described, "as goals centered around achieving a sense of personal pride and accomplishment." Satisfaction goals were described "as goals involved in achieving a sense of personal contentment or fulfillment. They are often related to the process and experience rather than just the outcome." Prior to each test, students completed a pre-test questionnaire about how they prepared for the test that week and how they thought they would perform. The pre-test questionnaire asked students to predict the score they expected to receive (Out of 100%, what percentage of the questions do you think you will answer correctly – i.e., what score do you think you will get on the test?), provide a score with which they would be satisfied (satisfaction goal – What score would you be satisfied with?), and a score with which they would be proud (pride goal- What score would you be proud of?) each on a scale of 0-100%.

Analytic Plan

Our goal was to determine if undergraduate students' satisfaction goals and pride goals for course tests would be better calibrated to their test performance than their predictions of test performance. Specifically, as our first hypothesis was that satisfaction goals would be better calibrated to test scores than the other two *predictors.* As such, our analytic plan was to first

aggregate (calculate means) of test scores, predicted test scores, pride goals, and satisfaction goals across the 11 exams. Next, we would conduct an RMANOVA with one within subjects factor and within subjects contrast, to determine if satisfaction goals and pride goals were different than test scores and if predicted test scores would be significantly different from satisfaction goals and pride goals.

Our next step would be to subtract aggregated test score means from each of the predictors to create difference scores. Differences scores would then be submitted to a similar repeated measure ANOVA to determine if the difference between both goal types and test scores would be different than the difference between test score predictions and test scores. Both analyses would allow us to determine if test score goals of both types would be better calibrated to actual test scores than test score predictions.

Results

Table 1 presents the means and standard deviations of test scores, predictions, satisfaction goals, and pride goals aggregated across the 11 tests. A within subjects (repeated measures) ANOVA using difference contrasts with variables entered in the order of mean aggregated test scores, satisfaction goals, predictions, and pride goals was significant, Wilk's $\lambda = .12$, $F(3,65) = 165.02$, $p < 001$. Within-subjects contrasts indicated that mean aggregated satisfaction goals were significantly higher than test scores, $F(1,67) = 77.11$, $p < 001$, predicted scores were higher than satisfaction goals, $F(1,67) = 120.64$, $p < 001$, and pride goals were higher than predictions, $F(1,67) = 410.77$, $p < 001$.

A more detailed analysis was conducted to examine these differences. For this analysis tests scores for each test were subtracted from the relevant satisfaction goals, predictions, and pride goals. These values were than averaged into mean satisfaction goal-test score differences,

69

Table 1. Aggregated Means and Standard Deviations Across Tests

Variable	Mean	Standard Deviation
Test Score	79.83	6.87
Satisfaction Goal	84.30	6.21
Prediction	85.67	6.10
Pride Goal	91.38	4.44

prediction-test score differences, and pride goal-test score differences. Figure 1 presents these difference scores. These difference scores were then subjected to a within subjects (repeated measures) ANOVA using simple contrasts with aggregated tests scores as the reference category contrasted against satisfaction goal-test score differences, prediction-test score differences, and pride goal-test score differences. Multivariate tests indicate significant differences among these difference scores, Wilk's $\lambda = .22$, $F(2,65) = 115.39$, $p < 001$. Within-subjects contrasts indicated that the mean satisfaction goal-test score difference ($M =4.48$, $SD = 4.12$) was significantly lower than the prediction-test score difference ($M =5.84$, $SD = 3.67$), $F(1,67) = 12.58$, $p < 001$, and the pride goal-test score differences ($M =11.56$, $SD = 4.47$), $F(1,67) = 215.77$, $p < 001$.
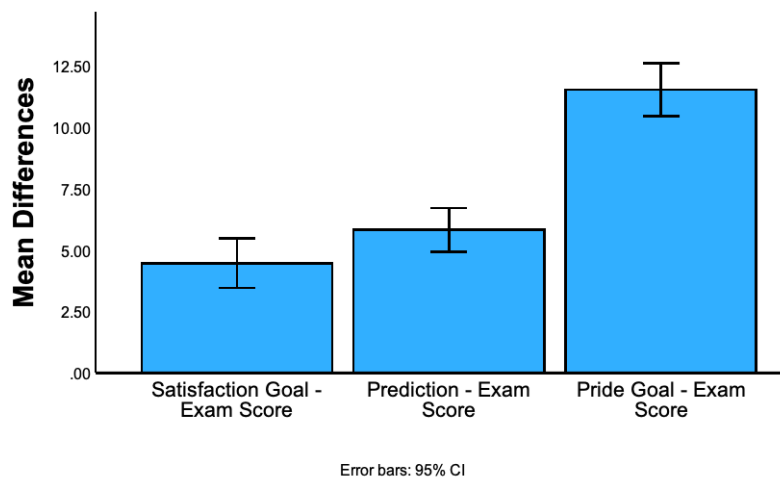

Figure 1. Mean Aggregated Difference Scores

Discussion

The overall results of our study replicate the findings that when students attempt to predict their test scores they are, for the most part, overconfident (Bol et al, 2005; Hacker et al, 2000; Isaacson & Was, 2010; Leman et al, 2023; Miller & Geraci, 2011). The results of our study also support our hypothesis that satisfaction goals would be better calibrated to test scores than test score predictions. We also hypothesized that pride goals, though slightly less well calibrated than satisfaction goals, would also be better calibrated than predictions; this was not the case. Students' pride goals were the least well calibrated to actual test scores.

 Research efforts to examine the ability of students to monitor their knowledge often demonstrate that students' knowledge monitoring (calibration) is highly inaccurate. When asked to predict their test scores, students, at all ranges of ability and performance, are overconfident, with the poorest performing students also being the least well calibrated. The current study found that if students are asked to state the test score with which they would be satisfied, they provide satisfaction goal scores that are significantly lower than their pride goals, and more importantly, significantly lower and better calibrated to their test score predictions. The reasons that satisfaction goals more accurately predict test performance are likely numerous. We propose three possible explanations.

The first explanation is that satisfaction goals accurately reflect the score students hope they will achieve based on their perceptions of preparedness. Satisfaction, as defined as the fulfillment of a want or need, may more accurately reflect the score for which students feel they prepared than their predicted score or their pride goal.

A second explanation is that in our study satisfaction goals acted like SOJs in previous studies (e.g., Al-Harthy et al, 2015). Recall that in previous research participants provided SOJs after their test score predictions. Put differently, in previous studies student-participants were asked to predict their test scores and then were asked how accurate they believed their predictions to be. In most cases, even poorly performing students recognized that their predictions may be in accurate (e.g., Dunlosky et al, 2005). In our study, student-participants made predictions regarding their test scores, then stated their satisfaction goals and pride goals in that order. Students may have recognized that their predictions may have been too high and thus their satisfaction goal may have reflected their awareness that predictions reflected overconfidence or that their predictions were possibly inaccurate.

Another explanation for the lower satisfaction goals is that satisfaction goals act as ego-protective goals (Boekaerts & Niemivirta, 2000). Put differently, students may set satisfaction goals that are lower than their predictions to mitigate the impact of not getting the score which they predicted. A student may predict a score of 90, but then state their satisfaction goal is a score of 85. "I think I am going to get a 90, but I would be ok with an 85."

In contrast to our predictions, pride goals were extremely higher than satisfaction goals and predicted scores. Pride goals might represent extrinsic motivations rather than internal motivations that might be reflected in satisfaction goals. From an achievement goal theory perspective (see Elliot, 2005; Urdan & Kaplan, 2020 for reviews) pride goals may reflect students' performance goals. Performance goals are based on measuring competence in comparison to others or normative comparisons (Was, 2006). Students may set pride goals that represent the score they think they would need to achieve to do well to perform well relative to other students.

A second explanation for the extremely high pride goals is the "wishful thinking" hypothesis (e.g., Lipko-Speed et al, 2019; Stipek et al, 1984). The wishful thinking hypothesis proposed by Stipek et al state that young children often fail to differentiate their outcome expectations from their desires. This may also be the case with undergraduates' pride goals.

Another perspective regarding the relationship between predicted scores, satisfaction goals, and pride goals is that of self as in important aspect of goal setting. Higgins (1987) made a distinction between the *ideal self* and the *ought self.* The ideal self represents the attributes a person would like to possess, and the ought self represents the attributes a person believes they should possess. Markus and Nurius (1986) proposed a similar idea suggesting that individuals compare their current view of themselves to *possible selves.* An interpretation of the current findings is that predictions and pride goals reflect goals based on possible selves. In the current study, pride goals may represent the ideal self. These high goals represent the scores students would like to achieve. Predicted test scores may represent ought selves in that these are the scores students feel they ought to achieve, and satisfaction goals may represent a more realistic self-concept about what students expect to achieve.

Future Directions

The results of the current study are informative yet there are a great number of questions left unanswered. The possible selves explanation of the current results needs to be explored. An individual differences study testing the relationship between the different possible selves and calibration between predictions, satisfaction goals, pride goals, and test scores is warranted. From and individual differences perspective, examining individual differences in goal orientations, metacognition, and study strategies in relationship to predictions, satisfaction goals, and pride would also be informative.

73

Future research should also investigate the relationship between satisfaction goals, calibration, and self-regulated learning. We can see many avenues that this research may travel. First, determine the direction of the relationship between satisfaction goals and self-regulated learning is important. Is it that students regulate their learning by setting satisfaction goals and then studying until they feel that they can achieve that goal? In the case of the current study, over the course of the semester and the several tests, students set satisfaction goals during the week and studied until they felt they could meet that performance goal. Alternatively, at the time of the test students may have retroactively reflected on their studying that week to set their satisfaction goal at the time of the test. If the former is correct, then working with students to set satisfaction goals for the week might help with self-regulation. If the later, satisfaction goals might be informative to instructors about students' self-perceptions of their test preparation and students' self-efficacy.

There is a plethora of research investigating whether training can improve students' metacognitive calibration (e.g., Al-Harthy et al, 2015; Miller & Geraci, 2011) with varying levels of success. Perhaps, helping students to better align their pride and satisfaction goals with test performance might better serve students self-efficacy and self-regulation. Studies investigating *goal training* may informative. Specifically, future studies could investigate the efficacy of interventions aimed at helping students recalibrate pride and satisfaction goals.

From a practical or application perspective, examining the influence of the different goals on performance, and the reciprocal relationship would also be important. A longitudinal investigation of the influence of SGs and PGs on performance and the influence performance on subsequent SGs and PGs would both be theoretically, as well as practically, important.

Limitations

74

It is warranted that we acknowledge some limitations to the current study. As mention in the Methods, this study was a part of a larger investigation regarding students' metacognition and self-regulation. The multiple tests administered throughout the semester may have influenced students predicted scores, as well as the pride and satisfaction goals. Courses in which there are fewer tests (e.g., a midterm and a final) we might expect students to set different pride and satisfaction goals. We also suspect that the multiple-choice format of the tests may have influenced students' prediction and goals. It is quite possible that other test formats, such as an essay or short answer format, would have elicited different predictions and goals.

A second limitation of our study is that there are many individual differences variables that might have acted as moderators. For example, students' goal orientations might affect their pride and satisfaction goals. Students who are more performance-approach in their goal orientations may set higher pride and satisfaction goals. Academic achievement might also impact students' pride and satisfaction goals. One might hypothesize that students with a history of academic success set higher satisfaction goals then students with a history of academic difficulties.

Our study is also limited in the generalizability to broader populations. The current study was conducted with an undergraduate course in educational psychology. This clearly limits the generalizability of our results. Future work should attempt to replicate our findings with other undergraduate courses or samples, and with students in K-12 settings.

Conclusions

Investigations of knowledge monitoring, as measured by calibration of test score predictions and actual test performance, has been an important aspect of the metacognitive and self-regulation literature. Previous investigations have consistently reported that students

demonstrate poor calibration with a tendency to be overconfident. The current investigation indicates that students' predictions of their test scores do not reflect the score students truly believe they will achieve on the test. Rather, students' satisfaction goals may be a more accurate reflection of the test scores students believe they will achieve.

References

Al-Harthy, I. S., Was, C. A., & Hassan, A. S. (2015). Poor performers are poor predictors of performance and they know it: Can they improve their prediction accuracy. *Journal of Global Research in Education and Social Science*, *4*(2), 93-100.

Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In *Handbook of self-regulation* (pp. 417-450). Academic Press.

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, *73,* 269-290.

Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. *The Journal of general psychology*, *132*(4), 335-346.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational behavior and human decision processes*, *105*(1), 98-121.

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. *Handbook of competence and motivation*, *16*(2005), 52-72.

García, T., & Pintrich, P. R. (1995, April). Assessing Students' Motivation and Learning Strategies: The Motivated Strategies for Learning Questionnaire. Paper presented at the annual meeting of the *American Educational Research Association*, San Francisco, CA,

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92,* 160–170. *Neuropsychology, and Cognition, 1,* 203-237.

Higgins, E. T. (1987). Self-discrepancy: a theory relating self and affect. *Psychological review*, *94*(3), 319.

Isaacson, R., & Fujita, F. (2006). Metacognitive knowledge monitoring and self-regulated learning. *Journal of the Scholarship of Teaching and Learning*, 39-55.

Isaacson, R. M., & Was, C. A. (2010). Believing You're Correct vs. Knowing You're Correct: A Significant difference? *The Researcher,23*(1), 1-12.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of personality and social psychology*, *82*(2), 180.

Leman, J., Kurinec, C., & Rowatt, W. (2023). Overconfident and unaware: Intellectual humility and the calibration of metacognition. *The Journal of Positive Psychology*, *18*(1), 178-196. doi.org/10.1080/17439760.2021.1975155

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology*, *103*(2), 152-166. doi.org/10.1016/j.jecp.2008.10.002

Markus, H., & Nurius, P. (1986). Possible selves. *American psychologist*, *41*(9), 954.

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: the influence of incentives and feedback on test predictions. *Metacognition Learning, 6,* 303-314.

Nederhand, M. L., Tabbers, H. K., De Bruin, A. B., & Rikers, R. M. (2021). Metacognitive awareness as measured by second-order judgements among university and secondary

school students. *Metacognition and Learning*, *16*, 1-14. https://doi.org/10.1007/s11409-020-09228-6

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1,* 159–179.

Pieschl, S. (2009). Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning*, *4*, 3-31. doi.org/10.1007/s11409-008-9030-4

Roeser, R. W., Eccles, J. S., & Strobel, K. R. (1998). Linking the study of schooling and mental health: Selected issues and empirical illustrations at.. *Educational psychologist*, *33*(4), 153-176.

Stipek, D. J., Roberts, T. A., & Sanborn, M. E. (1984). Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. *Child Development, 55,* 1983–1989.

Urdan, T., & Kaplan, A. (2020). The origins, evolution, and future directions of achievement goal theory. *Contemporary Educational Psychology*, *61*, 101862. doi.org/10.1016/j.cedpsych.2020.101862

Was, C. A. (2006) Academic achievement goal orientation: Taking another look. *Electronic Journal of Research in Educational Psychology, 4*(3), 529-550.

Weinstein, C. E., Palmer, D. R., & Schultz, A. C. (2002). LASSI. *User's Manual for those administering Learning and Study Strategies Inventory*.

Williams, W. M. (2004). Blissfully incompetent. *Psychological Science in the Public Interest*, *5*(3), i-ii.