

Why Grades Engender Performance-Avoidance Goals: The Mediating Role of Autonomous Motivation

Caroline Pulfrey
Université de Lausanne

Céline Buchs
Université de Genève

Fabrizio Butera
Université de Lausanne

Evaluation is an inescapable feature of academic life with regular grading and performance appraisals at school and at university. Although previous research has indicated that evaluation and grading in particular are likely to have a substantial impact on motivational processes, little attention has been paid to the relationship between grading and approach versus avoidance achievement goals, 2 fundamental concerns whenever evaluation is at stake. Three experiments, carried out in professional schools, revealed that expectation of a grade for a task, compared with no grade, consistently induced greater adoption of performance-avoidance, but not performance-approach, goals. Experiments 2 and 3 revealed that expectation of a grade, compared with no grade, consistently induced greater adoption of performance-avoidance goals even when grading was accompanied by a formative comment. Furthermore, Experiment 3 showed that reduced autonomous motivation measured after having completed a task for a grade versus no grade mediated the relationship between grading and adoption of performance-avoidance goals in a subsequent task. Results are discussed in the light of achievement goal and self-determination theory.

Keywords: grading, evaluation, performance-avoidance goals, autonomous motivation, self-determination

“You may turn over your papers and begin.” Those fateful words leave no one untouched, as performance evaluations, often normative, are a central feature of academic life (Ames, 1992). It is generally accepted that goal adoption is shaped by environmental contingencies (Elliot, 1999; see also Ames & Ames, 1981; Covington & Omelich, 1984) and that normative evaluations engender a preoccupation with performance concerns (Butler, 1987, 2006; Elliot, 1999; Elliot & Moller, 2003; Harter, 1978; Kluger & DeNisi, 1996; Nicholls, 1984). Hence, it should come as no surprise that in the case of achievement goals, generally defined as the purpose of achievement activity (Elliot & Sheldon, 1997), normative evaluations have been associated with the adoption of pretask performance-oriented goals (Butler, 2006), namely goals that concentrate attention on “the demonstration of competence relative to others” (Elliot, 1999, p. 169).

However, performance goals come in two kinds (Elliot, 1999; Elliot & Church, 1997): performance-approach goals (“focus on attaining normative competence”) and performance-avoidance goals (“focus on avoiding normative incompetence”; Elliot & Murayama, 2008, p. 614). This is an important distinction as, on the one hand, performance-approach goals have been positively linked to challenge construal, performance aspirations, performance, and self-esteem (Elliot & Moller, 2003). Performance-avoidance goals, on the other hand, have been systematically linked to anxiety, hopelessness, shame (Pekrun, Elliot, & Maier, 2006), low ability-related self-esteem, and pretask threat appraisals and negatively related to self-determination, perceptions of control, and feeling calm during evaluations (Elliot & McGregor, 2001; McGregor & Elliot, 2002). With this in mind, it is surprising that, as far as we know, no study has directly addressed the question of whether normative evaluations, more precisely classroom grades, increase pretask performance-approach or -avoidance goals. The first aim of this research was to answer this question, and the second aim, building on Elliot’s (1999) argument that environmental variables can indirectly influence goal adoption via the intermediary of underlying motives, was to explore the motivational mechanisms that might explain the grading–goal relationship.

In pursuit of these aims, we argue that students depend on grades (Covington & Müeller, 2001), as grades are determined by an external source and students have limited control over the quality of the grade, grading being explicitly or implicitly normative in nature (Deutsch, 1979; Thorndike, 1913) and the value of the work being decided by the external source. That is, students find themselves in a position of dependence vis-à-vis the social agents, teachers, or external bodies who control these valued

This article was published Online First June 20, 2011.

Caroline Pulfrey and Fabrizio Butera, Institut des Sciences Sociales, Université de Lausanne, Lausanne, Switzerland; Céline Buchs, Educational Sciences Section, Université de Genève, Geneva, Switzerland.

This work was supported by the Swiss National Science Foundation. We wish to thank Alain Clémence, Benoît Dompnier, Franziska Krings, Gabriel Mugny, Valérie-Anne Ryser, and Philippe Sarrazin for valuable comments on a previous version of this article. We also thank Michel Berney and Gill Wenger, directors of the Centre d’Enseignement Professionnel de Vevey and the Ecole Professionnelle du Chablais, as well as the teachers and students who participated in the experiments.

Correspondence concerning this article should be addressed to Fabrizio Butera, Institut des Sciences Sociales, Université de Lausanne, Vidy, CH 1015 Lausanne, Switzerland. E-mail: fabrizio.butera@unil.ch

resources. Dependence on those who distribute valued resources is the equivalent of powerlessness, and powerlessness has been associated with a basic inhibition or avoidance motivational orientation (Keltner, Gruenfeld, & Anderson, 2003).

We additionally argue that this dependence of students in graded evaluation situations reduces student autonomy (Ryan & Deci, 2000), as heteronomous forces—defined by Ryan and Deci (2006, p. 1562) as “regulation from outside the phenomenal self, by forces experienced as alien or pressuring, be they inner impulses or demands, or external contingencies of reward and punishment”—are perceived by the student to be controlling his or her behavior. Consequently, we hypothesize, first, that students facing a graded, as compared with a nongraded, evaluation will experience higher levels of performance-avoidance but not performance-approach goals and, second, that this relationship will be explained by the lower levels of autonomous motivation generated during prior graded evaluations. Finding answers to these questions and effectively increasing knowledge of the “functional significance” of grading (Ryan & Grolnick, 1986, p. 557) can provide a significant theoretical contribution to the motivation literature as well as a useful practical contribution for educators and assessors.

Grades and Performance Goals

In a typology of assessment feedback, Tunstall and Gipps (1996) described a continuum of assessment going from evaluative—in which judgments are made according to explicit or implicit norms and feedback is clearly positive or negative and is, directly or indirectly, linked to reward or punishment—to descriptive, in which feedback specifies attainment, constructs future achievement aims, and provides information on how to move toward them. Grades, a value-laden symbol indicating the relative quality of a performance that is a regular feature of school life (Pope, 2001), are positioned firmly on the evaluative side of this typology. The value of a grade reflects certain norms, and the grade attributed reflects the degree to which normatively determined standards have or have not been attained.

In addition to this, a characteristic of grading is, as Thorndike (1913, p. 288) remarked, “its relativity.” That is, grades make assigned level recognition easy by their normative aspect (relativity), which encourages social comparison and renders hierarchy salient by heightening the importance of one’s normative standing. Thus, a grade is considered positive or negative not only according to the degree to which content-related standards are met but also according to performance relative to that of peers. Criterion-based grading aims to reduce this overtly comparative nature of grades by focusing attention on reaching set standards. However, as Elliot and Moller (2003) argued, social comparison of grades provides additional and more accurately diagnostic competence information than that provided by a task-focused evaluation standard and, as such, is commonly used by students to situate themselves even in classrooms that promote criterion-based grading.

With this in mind, it is probably not surprising that the relatively long tradition of achievement goal theory (Ames, 1992; Butler, 1987; Nicholls, 1984) has consistently argued that grades engender a performance focus, as performance achievement goals are generically considered as normative in content (Elliot & Murayama,

2008). Indeed, in 1984 Nicholls argued that competition on valued tasks and on tasks that are presented as tests of valued skills should all induce an ego orientation (i.e., performance goals), a focus on the protection of self-worth. Later Ames (1992) claimed that the normative aspect of grading is responsible for generating the ego-oriented focus on ability. These arguments are further backed up by the empirical work of Butler and Nisan (1986), which showed that attributing effort in the task to the perceived importance of success and the desire to avoid lack of achievement was greater in grades than in formative comment conditions. More specifically, attributions of effort to the desire to perform better than others and not to perform worse than others were also higher in grades than in conditions providing solely feedback with comments (Butler, 1987).

In line with this, Kluger and DeNisi (1996), in their presentation of feedback intervention theory, used grading as an example of an external intervention likely to render metatask processes concerning self-goals, such as performance goals, salient at the expense of task-related goals. According to feedback intervention theory, behavior is regulated by a comparison of feedback with goals or standards, which are organized hierarchically, with metatask processes concerning self-goals at the top, task-motivational processes concerning task-related goals in the middle, and task-learning processes concerning physical specific action goals at the bottom. Graded assessment is a feedback intervention from an external source with potentially serious implications and consequently moves individuals from the middle to the higher level of the goal hierarchy by putting the spotlight on metatask motivational, cognitive, and affective processes relating the focal task to higher order self-focused goals (Kluger & DeNisi, 1996).

Importantly, both Elliot (1999) and Pintrich (2000) have argued that strong situational cues or contexts can override the effects of chronic goal orientation. In the case of graded assessments, when is the situational cue most salient? According to Harackiewicz, Manderlink, and Sansone (1992), when a performance-contingent reward such as a grade is attached to performance, performance evaluation is anticipated and can arouse motivational and emotional consequences in the individual prior to engaging in a task. Thus, pretask performance goals would seem to be a fruitful state goal area to explore, particularly as pretask performance goal adoption has important consequences on task performance and related outcomes (Elliot & McGregor, 2001; McGregor & Elliot, 2002).

Grades and Performance-Approach Versus -Avoidance Goals

However, do grades increase approach or avoidance performance goals? According to Elliot and Moller (2003, p. 346), it is “the normative evaluation structures in the school and classroom environment that tend to evoke performance-approach goals,” and this viewpoint echoes an earlier comment of Deutsch (1979), who posited that grades have a manifest function as a tool of socialization used to prepare students for the competitive, meritocratic world of adult professional life. Indeed, it would seem possible that the competition involved in normative assessment should inspire competitive, performance-approach goals.

Yet if we concretely compare the motivational and behavioral outcomes of grade-based evaluation with those of performance-

approach and performance-avoidance goals, grading practices seem indeed to have more in common with the latter goal. First, although grading has been shown to provoke challenge avoidance (Harter, 1978), as have performance-avoidance goals (McGregor & Elliot, 2002), performance-approach goals are related with challenge construal (Elliot & Moller, 2003). Second, although grading has been associated with reductions in performance (Butler, 1988; Harter, 1978), as have performance-avoidance goals (Elliot & McGregor, 2001), performance-approach goals have, in a majority of cases, been associated with performance increase (Elliot & McGregor, 2001; Elliot & Moller, 2003). Third, grades (Butler, 1987, 1988; Butler & Nisan, 1986; Harter, 1978) and performance-avoidance goals (McGregor & Elliot, 2002) consistently seem to reduce intrinsic motivation, but performance-approach goals seem to have either a positive or null relationship with it (Elliot & Moller, 2003).

Why is this? Normative evaluation is characterized by the attribution of a value to a performance by an external source, teacher, or assessment body (Elliot & Moller, 2003). In the case of the grade, this value carries significant social and esteem-based present and future consequences (Darnon, Dompnier, Delmas, Pulfrey, & Butera, 2009). As Deutsch (1979, p. 396) said, "Grades, like money, have external value." Not surprisingly, therefore, grades have been shown, in a study carried out with college students, to be perceived by the majority of students as a particularly powerful form of reward and/or punishment, tokens that control one's future, "potent symbols, almost objects, that influence how one feels today and what one will be able to do tomorrow" (Goulden & Griffin, 1997, p. 34). As such, students depend heavily on the grades they receive. This dependence, allied with the fact that the student has limited control over the quality of the grade, owing to the implicitly or explicitly zero-sum nature of normative evaluation (Covington, 2000) as well as limited control over the timing and nature of the feedback process (Elliot & Moller, 2003), means that students receiving grades are in a position of relative powerlessness in the grade attribution process, depending in many cases on another individual, the teacher.

Power, defined as relative ability to "modify others' states by providing or withholding resources or administering punishments" (Keltner et al., 2003, p. 265), is squarely held by the teacher or assessment body. Relating power to the approach versus inhibition or avoidance distinction, Keltner et al. (2003) used a number of research findings to demonstrate that powerlessness is related to inhibition or avoidance tendencies such as negative affect, heightened sensitivity to threat and punishment, and behavioral inhibition. Elliot and Covington (2001) made the distinction between approach versus avoidance and *can* versus *may*, respectively. This reflects nicely the distinction between power and powerlessness. On the basis on these findings, we argue that having a performance evaluated with a grade is likely to increase the level of performance-avoidance goals, rather than performance-approach goals, precisely because the individual being evaluated is in a position of dependence and powerlessness. In sum, our first general hypothesis is that normative evaluation, in the form of a grade, will increase pretask performance-avoidance goals but not performance-approach goals.

Autonomy: The Why of the Grade-Performance-Avoidance Goal Relationship

If the key factor in being graded for a piece of work is indeed the unfavorable power balance faced by the student, how does this translate into motivational tendencies within the individual? We argue that autonomous motivation or self-determination is the missing piece in the puzzle, that engaging in normative assessment in a situation of powerlessness, dependence, and hence a lack of autonomy is what renders performance-avoidance as opposed to performance-approach goals salient in the minds of students about to undertake a graded task. In doing so, we take up the fine distinction made by self-determination theory (Deci & Ryan, 1985) and in particular the perceived-locus-of-causality framework, which lies at the heart of self-determination theory (Ryan & Connell, 1989). This framework organizes reasons for acting into a range going from controlled to autonomous. Controlled motivation is constituted of, first, extrinsic motivation with external regulation, purely external reasons for performing an action, such as fear of punishment or compliance with rules; and second, extrinsic motivation with introjected regulation, namely acting out of shame, guilt, or desire for socially derived ego enhancement. Autonomous motivation, on the other hand, is composed of, first, extrinsic motivation with identified regulation, behavior driven by an individual's values or goals; and second, intrinsic motivation, behavior energized by interest or enjoyment (Ryan & Connell, 1989).

Grading and Autonomous Motivation

Deci and Ryan (1987) argued that what determines the extent to which motivation is autonomous is not simply the goal pursued but also the degree to which one is or feels under pressure to pursue that goal. Elliot and Moller (2003, p. 345) took up a similar idea when they talked about normative strivings being likely to be "sullied by" less adaptive motivational concerns such as self-protection when "they are pushed from without (i.e., suggested or encouraged by others or the structure of the evaluative environment) rather than being allowed to emerge naturally from within." DeCharms (1968) indeed argued that being induced to participate in ego-involving activities creates a situation in which failure to attain goals represents a consequential threat to self-esteem, and therefore ego-involving inductions are perceived as reducing autonomy, that is, they are characterized by an external locus of causality. Following in his footsteps, Ryan (1982), citing the very example of evaluation, posited that the more one perceives communications from others as pressure to attain a specific outcome, the less likely one is to experience autonomous motivation for that activity.

Relating this framework to grading, one can discern a distinct potential contradiction between the state of powerlessness and dependence that characterizes the student facing a graded evaluation and the characteristics of autonomous motivation described above. Deutsch (1979, p. 393) summed this up, saying, "Grades, by serving as a motivating value dispensed by a controlling authority, enormously facilitate the socialization of the young without resort to excessive violence or bribery by those having authority over students."

Consequently, it is not perhaps surprising that grading has often been argued to reduce feelings of autonomy in students (Butler &

Nisan, 1986; Deci & Ryan, 1985; Grolnick & Ryan, 1987; Harter, 1978; Ryan, 1982; Ryan, Connell, & Deci, 1985). As Butler and Nisan (1986, p. 210) commented, "The normative grades prevalent in many schools seem to provide a striking example of feedback in which control rather than information is salient."

Autonomous Motivation and Performance-Avoidance Goals

Autonomous motivation has been presented by Elliot and McGregor (2001) as an essentially appetitive urge and has been found to negatively predict avoidance goals. In the same vein, Covington and Müeller (2001) posited that intrinsic and extrinsic motivation correspond to approach and avoidance tendencies, respectively. Furthermore, reduced autonomous motivation has been argued to relate to negative affect (Smith, Sinclair, & Chapman, 2002), to diminish effects of positive feedback on subsequent behavior (Boggiano & Barrett, 1985), and to generate classroom disaffection and lower levels of achievement (Eccles, 1998). These fundamentally negative consequences of reducing autonomous motivation allied with the image-focused pressure to succeed and reduced intrinsic motivation typical of graded contingencies (Butler & Nisan, 1986; Harter, 1978) imply that a reduction in autonomous motivation would indeed generate subsequent pretask performance goals that are negative, that is, performance-avoidance goals. Further support for this is provided by Elliot & McGregor's (2001) finding that autonomous motivation is negatively related to performance-avoidance goals and unrelated to performance-approach goals.

Why? We posit that grade evaluations generate pressure to engage in an activity and to do well in it in order to attain the value and limited resource of a good grade. Furthermore, in a graded evaluation, much is at stake, as failure to succeed may not simply leave the individual at the starting block but worse off than before, with a lower grade point average (GPA), loss of self-esteem and face, feelings of guilt, and potential retributions from powerful others. Prospect theory (Kahneman, Knetsch, & Thaler, 1991) teaches that as stakes increase, aversion to loss increases exponentially, considerably more than equivalent attraction to gain. Therefore, it is indeed possible that avoiding substandard performance might well be uppermost in students' minds as they embark upon graded tasks.

Building on the arguments we have laid out in these last two sections, we posit that accomplishing a graded activity erodes autonomous motivation, and this erosion is responsible for the adoption of performance-avoidance goals. Consequently, our second general hypothesis is that a reduction in autonomous motivation experienced during a graded activity should explain why grading increases subsequent pretask performance-avoidance goals.

The Role of Perceived Competence

A question remains unanswered. Competence lies at the heart of achievement goal theory. Achievement goals are defined as the purpose of competence-relevant behavior, and performance evaluations are used to determine absolute, intrapersonal, or normative competence (Elliot & McGregor, 2001). Yet the question as to the role perceived competence plays in the goal adoption process is

still not solved. Although Elliott and Dweck (1988) presented perceived competence as a potential moderator of the relationship between goals and a range of learning outcomes, more recently, Moller and Elliot (2006), citing the failure of Cury, Elliot, Da Fonseca, and Moller (2006) to find a moderating role of perceived competence in the achievement goal to performance relationship, tended to favor the potential of perceived competence to act as an antecedent of achievement goal orientation in studying.

Furthermore, Elliot (1999) argued that competence perceptions are but one among a whole series of potential intrapsychic and environmental goal predictors, adding that if the achievement-setting cues are sufficiently dominant, they have the potential to determine goal adoption directly via "situation-specific concerns" (p. 176). This issue would seem particularly relevant in the case of grading, as grading and normative evaluations in general are strong environmental cues that leave no individual, however high-achieving, untouched (Pope, 2001). Results of a key study by Butler (1988) back this up, revealing that, in a comparison of three classroom feedback conditions—grade-only feedback, grade-with-comment feedback, and formative-comment-only feedback—both high-achieving (the top 25% GPA) and low-achieving (the bottom 25% GPA) fifth and sixth graders suffered deficits in performance and ongoing motivation in both graded conditions, compared with the formative-comment-only condition. In sum, perceived competence may intervene in goal adoption, but its effects may also be overridden by situation-specific concerns such as grades. Thus, perceived competence is included in two of the reported studies in order to test whether it is a moderator of the predicted effect of grading on performance-avoidance goals.

Hypotheses and Overview

It would seem that carrying out an activity for a grade will be associated with higher levels of pretask performance-avoidance goals, compared with carrying out an activity that will not be graded, and that one explanatory mechanism for this relationship could be the reduction in autonomous motivation experienced in graded activities. In the light of these reflections, three experiments were carried out to test the following hypotheses. First, we hypothesize that in conditions in which a grade is anticipated compared with conditions in which a grade is not anticipated, pretask performance-avoidance goal adoption should increase, a difference that should not appear for pretask performance-approach goal adoption. Second, we hypothesize that the relationship between anticipating receiving a grade and pretask performance-avoidance goals will be mediated by the lower levels of autonomous motivation experienced in previously graded assessment conditions in which no graded assessment was present.

Experiment 1 compared graded and nongraded conditions and was designed to test the first hypothesis. Our second experiment was designed to replicate Experiment 1 with this time a more stringent assessment-focused comparison. Indeed, Experiment 1 confounds the absence of grades with the absence of assessment; in Experiment 2, we compared the grade-only, comment-only, and grade-with-comment conditions. In addition, as discussed above, it is possible that perceived competence intervenes in goal adoption, but also that its effects are overridden by situation-specific concerns such as grades; thus, we tested the potential moderating effects of perceived competence in class exercises. In Experiment

3, we first replicated the experimental design used in Experiment 2 to test Hypothesis 1, using this time perceived competence in the specific task to be engaged in. Then, using a longitudinal method, we tested our second hypothesis that the lower levels of autonomous motivation experienced in a graded assessment setting as opposed to a nongraded assessment setting would act as a mediator in the relationship between graded versus nongraded assessment conditions and the degree of subsequent endorsement of performance-avoidance goals prior to engaging in a task.

All three experiments were field experiments carried out within the classroom context of Swiss professional schools. Different sets of students participated in each of the three studies. In all experiments, the procedure and materials were approved by the school authorities, administrators, and teachers. Students were given the option not to participate in the experiment if they wished, although they were required to remain in the class and carry out normal class activities. All students freely consented to participate. Students in these schools are used to receiving regular graded assessments to carry out in class time. Feedback provided by the teacher usually comprises a grade out of 6 with 4 as the pass mark and a limited number of written annotations, mainly ticks or crosses. In all experiments, classes were randomly assigned to each condition.

Experiment 1

The aim of our first experiment was to test our first hypothesis, that in conditions in which a grade is anticipated compared with conditions in which a grade is not anticipated, pretask performance-avoidance goal adoption should be higher, a difference that should not appear for pretask performance-approach goal adoption.

Method

Participants. One hundred and fifteen students attending a Swiss professional school (akin to a technical college), with a mean age of 19.5 years ($SD = 3.43$; range: 15–30, with seven participants over 24), participated in this experiment. The sample consisted of 42 male and 73 female students in eight classes, with four classes allocated to each of two conditions: graded and nongraded task. Although there were significantly more female than male students in our sample, $\chi^2(1, N = 115) = 8.36, p < .01$, there were no significant differences between numbers of male and female students in each condition, $\chi^2(1, N = 114) = 0.01, ns$.

Procedure. The experiment was carried out during a regular English as a foreign language (EFL) class and was presented to the students with the collaboration of the class teacher as a pretask survey designed to find out their opinions of the course material. At the beginning of the class, a researcher entered the classroom with the class teacher and asked the students whether, prior to their carrying out an anticipated individual activity (an English listening comprehension) organized and presented by their teacher, they would be willing to fill in a questionnaire about the upcoming activity. All students agreed to fill in the questionnaire. The subject teacher then announced the activity and the conditions of assessment. The activity was a listening comprehension, a habitual exercise in EFL classes that is regularly carried out either for a grade or as a simple ungraded classroom exercise. In the graded condition, the teacher informed the students that the listening

comprehension would be assessed with a grade. In the professional school system, as in the high school system, graded in-class exercises are regular activities for which the result systematically counts toward the semester subject GPA. In a language class the end-of-semester subject GPA will consequently be the average of up to 20 or so classroom assessments. In the nongraded condition, the teacher presented the activity and specified that no grades were involved. The students filled in the survey, which was collected by the researcher, who then left the classroom, and the class then commenced the class activity. Performance goal measures were collected just prior to task commencement, as environmentally induced goals are likely to be at their most salient and influential at this point (see Elliot, 1999; McGregor & Elliot, 2002). As agreed with the teacher, the task was neither collected by the experimenter nor assessed by the teacher. Students were fully debriefed and informed of the purpose of the study.

Measures. The students' performance goals were measured with scales from a task-level version of the French translation (Darnon & Butera, 2005) of the 2×2 Achievement Goal Questionnaire (Elliot & McGregor, 2001). The two scales, ranging from 1 (*not at all*) to 7 (*completely*), were performance approach (e.g., "It is important for me to do better than others"; $\alpha = .87$) and performance avoidance (e.g., "My goal in this exercise is to avoid performing poorly"; $\alpha = .71$). Of note is the fact that the three-item scale of performance avoidance produced an initial alpha of .60, and only when the item "My fear of performing poorly in this exercise is what motivates me" was removed did the alpha rise to .71, as in Darnon, Harackiewicz, Butera, Mugny, and Quiazade (2007). Consequently, a two-item version of this scale was retained for subsequent analyses. Descriptive statistics are presented in Table 1.

Results

To test the effects of anticipating a grade on pretask performance-avoidance goals, we contrasted the condition in which students anticipated receiving a grade (coded $-.5$) with the condition in which students did not anticipate receiving a grade (coded $.5$). As our experimental conditions were made up of classes of students, we constructed a mixed-effects model in SPSS to test our first hypothesis while controlling for random effects from classes used in the experimental conditions. Analyses were carried out in two stages. In the first stage, restricted maximum likelihood estimation (RMLE) was used in two models: an "unconditional" model with no predictors, which tested the contribution of class to variance in performance-avoidance goals and measured intraclass correlations (Albright & Marinova, 2010), and a second model, which examined the effect of class once our fixed effect, that of anticipating a grade versus no grade for a piece of work, was included in the model. RMLE was used at this stage to control as accurately as possible for class effects, as full maximum likelihood estimation (MLE) is biased toward lower parameter estimates in samples in which the number of Level 2 groups is low (Bressoux, 2008; Garson, 2010). In the second stage, we reran the same two models using MLE. This enabled us to use the measure of relative deviance of alternative models to assess relative model fit to the data and to compare fixed effects. RMLE, which includes only variance components, not regression coefficients, as is the case

Table 1
Experiment 1 ($N = 115$): Descriptive Statistics and Zero-Order Correlations Among Variables

Variable	Grade			No grade			1	2
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range		
1. Performance-approach goals	3.03	1.48	1–6	3.06	1.43	1–6	—	.31**
2. Performance-avoidance goals	5.27	1.14	2–7	4.62	1.30	1–7	—	—

** $p < .01$.

with MLE, cannot be used for model comparison of fixed effects (Bressoux, 2008; Garson, 2010).

Stage 1: Class variance. Results of our first unconditional model, with class entered as a random effect, indicated a nonsignificant contribution to performance-avoidance goal adoption by class (Wald $Z = 1.20$, ns) and between-class residual ($\sigma^2 = 1.39$, $SE = 0.19$, $p < .001$) over 8 times that of the between-class variance ($\tau^2 = 0.17$, $SE = 0.14$), implying that the majority of variance was unexplained in this model. However, the intraclass correlation coefficient (ICC) indicated that when the experimental conditions in which classes were nested were not included in the model, approximately 11% of the variance in performance-avoidance goal adoption was explained by the class. Consequently, it was deemed prudent to include class-level random effects in subsequent models (see Singer, 1998).

In Model 2, our fixed effect was added, contrasting the condition in which students anticipated receiving a grade (coded $-.5$) with the condition in which students did not anticipate receiving a grade (coded $.5$). Class was included as a random effect to control for the contribution of class to pretask performance-avoidance goal adoption. Results of Model 2 indicate that between-class variance ($\tau^2 = 0.07$, $SE = 0.11$, ns) in Model 2 was 0.1 less than that of Model 1 and the between-class residual ($\sigma^2 = 1.40$, $SE = 0.19$, $p < .001$) was now nearly 20 times that of the between-class variance. The ICC dropped from 11% in Model 1 to 5%. This would seem to imply that variance due to class did not have any significant impact on pretask performance-avoidance goal adoption and that variance in pretask performance-avoidance goal adoption not arising from student class was of a considerably greater magnitude than any that did arise from student class.

Stage 2: Fixed effects and model comparison. In Stage 2, the focus was on testing the contribution of our fixed effect, that of receiving a grade versus no grade on performance-avoidance goal adoption and on testing the relative fit of the unconditional Model 1 with the mixed Model 2 to the data. Our fixed-effect contrast, the grade contrast, had a significant effect on pretask performance-avoidance goals ($\gamma = -0.65$, $SE = 0.24$, $p < .05$), with students reporting higher levels of performance-avoidance goals in the grade condition ($M = 5.27$, $SD = 1.14$), compared with the no-grade condition ($M = 4.62$, $SD = 1.30$). Deviance in Model 1, the unconditional model, was 371.15. Deviance in Model 2, the model including fixed as well as random effects, dropped to 366.70, a significant difference of 4.45, $\chi^2(1, N = 115) = 4.45$, $p < .05$, indicating that Model 2 with fixed effects provided a significantly improved fit to the data. Results are reported in Table 2.

When the same Stage 1 and 2 analyses were run on pretask performance-approach goals, Stage 1 analyses produced an ICC of 0%. Stage 2 analyses showed that our fixed-effect contrast, the

grade contrast, had no effect on pretask performance-approach goals ($\gamma = -0.03$, $SE = 0.27$, ns), which did not differ between graded ($M = 3.03$, $SD = 1.48$) and nongraded conditions ($M = 3.06$, $SD = 1.43$).

Discussion

These results provide preliminary evidence for our first hypothesis, showing that the anticipation of a grade for an activity increased pretask performance-avoidance goals when compared with a nongraded activity, an effect that did not manifest itself in the case of performance-approach goals.

Worthy of note in this first experimental study is the initially low performance-avoidance goal alpha and the fact that only two of the three habitual performance-avoidance goal items cohered, as also found in recent research by Darnon et al. (2007). It is interesting that the issue of the third item, "My fear of performing poorly in this exercise [class] is what motivates me," has since been addressed by Elliot and Murayama (2008) in a rehabilitation

Table 2
Experiment 1: Multilevel Model Testing the Impact of Graded Versus Nongraded Assessment on Students' Pretask Performance-Avoidance Goals, Controlling for Effects of Student Class

Variable	Model 1		Model 2	
	Estimate	<i>SE</i>	Estimate	<i>SE</i>
Fixed effects ^a				
Intercept	4.97*** ^b	0.17	4.99***	0.12
Grade contrast			-0.65*	0.24
Random effects ^c				
Intercept	0.17 ^d	0.14	0.07	0.11
Residual	1.39*** ^e	0.19	1.40***	0.19
Model fit statistics ^d				
Deviance	371.15		366.70	
AIC	377.15		374.70	
BIC	385.39		385.67	
ICC	.11		.05	

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; ICC = intraclass correlation coefficient.

^a Using maximum likelihood estimation. ^b Overall sample level of pretask performance-avoidance goal adoption. ^c Using restricted maximum likelihood estimation. ^d Variance component corresponding to random intercept. ^e Class-level error.

* $p < .05$. *** $p < .001$.

of the achievement goal scales. According to these authors, this particular item stands out from the others in that it collapses the goal and the motivation underlying the goal, in this case fear of failure. This is a problem, as, in the 2×2 goal framework (Elliot & McGregor, 2001), a goal is seen as “a cognitively represented aim,” which is separate from “the reason or reasons why the person is pursuing the aim” (Elliot & Murayama, 2008, p. 614). This distinction is particularly relevant in the comparison between pretask goals adopted in graded and nongraded assessments. The fact that the average student, when faced with a habitual classroom assessment, might reasonably have a cognitive objective of not doing badly without an underlying fear-of-failure motive, could well explain the improved scale reliability without this item. Thus, the use of this two-item score seems relevant to the present purpose; however, to be on the safe side, we sought a replication of the effect of grading on performance-avoidance goals in a further experiment.

In replicating Experiment 1, we clarified an important issue, related to the manipulation of grading. Indeed, in Experiment 1 grades and assessment were confounded. That is, it is possible that the observed effects are due to the presence and absence of grades, or to the presence and absence of teacher assessment. To solve this problem, we built on the literature on formative assessment, which contrasts graded assessments with individualized task-based feedback in the form of comments on student work (e.g., Black & Wiliam, 1998; Crooks, 1988; Natriello, 1987; Sadler, 1989). Would the observed results still hold if we compared graded versus nongraded teacher assessment? To achieve this, in Experiment 2 we let all conditions involve teacher feedback. However, the presence and absence of grades were manipulated in three conditions: the first one being the receipt of a grade alone, the second being the receipt of a grade accompanied by a formative comment, and the third being a formative comment unaccompanied by any grade. This would provide a more stringent test of the hypothesized grade to performance-avoidance relationship, as teacher feedback, even comment based, could be perceived as potentially threatening (Koka & Hein, 2003; Peters, 1978).

A second aim of Experiment 2 was to test the potential moderating impact of perceived competence on the grade to pretask performance-avoidance goal adoption relationship. This was accomplished by including a measure of self-efficacy experienced in the exercises carried out within the context of the subject, in this case EFL.

The third question that needs to be dealt with is whether being faced with an unanticipated grading situation, as occurs with pop quizzes, might exacerbate the performance-avoidance goal adoption effect. Although EFL comprehension exercises tend to test general linguistic progression rather than specific, focused content, it could be argued that being mentally prepared for a graded exercise might influence pretask goal adoption. Consequently, in both Experiments 2 and 3, subject teachers informed students of the upcoming activity, be it accompanied by a grade, comment-based feedback, or both, during the previous week's class.

Experiment 2

Consequently, our second experiment aimed to conceptually replicate the results of Experiment 1 with a more stringent experimental design. The specific hypothesis was that both the grade-

only and grade-with-comment conditions would elicit a higher level of performance-avoidance endorsement than the comment-only condition, a difference that should not appear for performance-approach goals. Moreover, we wished to test the potential moderating role of task-level perceived competence through a measure of self-efficacy in the exercises carried out within the class.

Method

Participants. One hundred and thirty students attending a Swiss professional school participated in this experiment. Eight participants had uncommon studentized deleted residual on relevant measures and were dropped from the analyses (Judd & McClelland, 1989). The final sample consisted of 122 students, 80 male and 42 female students, with a mean age of 17.23 years ($SD = 1.56$; range: 15–23), in 12 classes. Four classes were allocated to each of three conditions—grade only, comment only, and grade with comment—giving 42 students in the grade-only condition, 38 in the comment-only condition, and 42 in the grade-with-comment condition.

Procedure. The experimental procedure was the same as that used in Experiment 1, except in three respects: First, the task and assessment conditions were not a surprise, as the teacher had announced them the previous week in class; second, a measure of task-level self-efficacy was taken prior to the in-class announcement of the assessment conditions; and third, grading was now manipulated through the three conditions presented above: grade only, grade with comment, and comment only. In the grade-only condition, students' work was assessed with a grade that would count toward the semester subject GPA. In the grade-with-comment condition, it was assessed with a grade that would count toward the semester subject GPA and an individualized comment. In the comment-only condition, it was assessed with an individualized comment, specifying areas of the work mastered and giving indications how to improve areas of the work further (e.g., “You have shown you have understood the main theme of the discussion. Now try to distinguish more precisely the viewpoint of each speaker. . . . Listen out for key words in each voice”; see Butler, 1987, 1988). In both conditions featuring comment-based feedback, the principle of the comment-based feedback was explained to the students with examples given before the achievement goal questionnaire was filled in.

Measures. Measures of participants' performance-related achievement goals and perceived competence in class exercises were taken.

Achievement goals. The students' performance goals were measured with the same achievement goal scale as in Experiment 1. Dimension reliability was once again fairly satisfactory (performance approach, $\alpha = .84$; performance avoidance, $\alpha = .64$). Of note once again is the fact that the three-item scale of performance avoidance produced an initial alpha of .55, and only when the item “My fear of performing poorly in this exercise is what motivates me” was removed did the alpha rise to .64. Consequently, a two-item version of this scale was retained for subsequent analyses.

Perceived competence in class exercises (task-level self-efficacy). The students' perceived competence in class exercises was measured with class exercise-based questions taken from

Midgley et al.'s (1997) Patterns of Adaptive Learning Survey scale of self-efficacy. The scale was composed of three items (e.g., "I'm certain I can understand even the hardest exercises we do in this class"; $\alpha = .83$). Items were translated into French and then back-translated blind in order to ensure compatibility in meaning between languages. As the measure of task-level self-efficacy was taken after the announcement the previous week by the teacher of task and assessment conditions, a one-way analysis of variance was used to test for differences in task-level self-efficacy between the three conditions. Results showed no significant difference between the different assessment conditions, $F(2, 130) = 1.62, ns$. Descriptive statistics are presented in Table 3.

Results

To test our more stringent comparison between graded and nongraded assessment, we created two orthogonal contrasts using Helmert coding: Contrast 1, the grade contrast, with the comment-only feedback condition coded 2 and the grade-with-comment and grade-only conditions coded -1; and Contrast 2, the comment contrast, which tested the impact of grade with comment versus grade alone, with the comment-only condition coded 0, the grade-and-comment condition coded -1, and the grade-only condition coded 1. A mixed-effects model was constructed in SPSS to test our two hypotheses, as this enabled us to control for random effects from classes used in the experimental conditions. As before, analyses were carried out in two stages. In the first stage, RMLE was used in the two models to assess variance due to class as opposed to fixed effects. In the second stage, full MLE was used to assess the relative impact of the fixed effects and to compare model fit.

Stage 1: Class variance. Results of our first unconditional model, with class entered as a random effect, indicated a nonsignificant contribution to performance-avoidance goal adoption by class (Wald $Z = 1.30, ns$) and between-class residual ($\sigma^2 = 1.19, SE = 0.13, p < .001$) over 8 times that of the between-class variance ($\tau^2 = 0.10, SE = 0.08$), implying that the majority of variance was unexplained in this model. The ICC indicated that when the experimental conditions in which classes were nested were not included in the model, approximately 6% of the variance in performance-avoidance goal adoption was explained by the class.

In Model 2, fixed effects were added, including the grade contrast and the comment contrast, in order to test the impact of the two contrasts on pretask performance-avoidance goals. Task-level self-efficacy was also included as a potential moderator of the impact of each contrast on pretask performance-avoidance goal

adoption, and thus interactions between task-level self-efficacy and the two contrasts were also added to the model. This gave five fixed effects. Class was included as a random effect to control for the contribution of class to pretask performance-avoidance goal adoption. Results of Model 2 indicate that between-class variance ($\tau^2 = 0.00, SE = 0.00$) in Model 2 was 0.1 less than that of Model 1 and, consequently, that once again variance due to class did not have any significant impact on pretask performance-avoidance goal adoption.

Stage 2: Fixed effects and model comparison. In Stage 2, the focus was on testing the contribution of our fixed effects on performance-avoidance goal adoption and on testing the relative fit of the unconditional Model 1 with the mixed Model 2 to the data. Results of Stage 2 analyses indicate that the grade contrast had a significant effect on pretask performance-avoidance goals ($\gamma = -0.28, SE = 0.07, p < .001$), with students reporting higher levels of performance-avoidance goals in the grade conditions ($M = 5.73, SD = 0.97$), compared with the comment-only condition ($M = 4.95, SD = 1.30$). No other fixed effects reached significance. Deviance in Model 1, the unconditional model, was 557.29. Deviance in Model 2, the model including fixed as well as random effects, dropped to 383.71, a significant difference of 173.58, $\chi^2(5, N = 130) = 173.58, p < .001$, indicating that Model 2 with fixed effects provided a significantly improved fit to the data. Results are reported in Table 4.

When the same Stage 1 and 2 analyses were run on pretask performance-approach goals, Stage 1 analyses produced an ICC of 4.5%, variance due to class being nonsignificant ($\gamma = 0.09, SE = 0.10, ns$, Wald $Z = 0.94, ns$). Stage 2 analyses showed that our fixed-effect contrast, the grade contrast, had no effect on pretask performance-approach goals ($\gamma = -0.04, SE = 0.12, ns$), which remained strictly similar in graded ($M = 3.72, SD = 1.40$) versus nongraded conditions ($M = 3.63, SD = 1.57$). No other fixed effects reached significance.

Discussion

The results of this experiment provide evidence that the anticipation of a grade for an activity increases pretask performance-avoidance goals when compared with nongraded comment-based assessment, thus extending the results of Experiment 1 to a specific comparison between the anticipation of graded and nongraded teacher feedback. Parallel to Butler's (1987) work on the grade-intrinsic motivation relationship, this result indicates that performance-avoidance goals are a result of graded assessment in particular as opposed to assessment in general: Assessment based uniquely on comments produced lower levels of performance

Table 3
Experiment 2 ($N = 122$): Descriptive Statistics and Zero-Order Correlations Among Variables

Variable	Grade only			Grade and comment			Comment only			1	2	3
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range			
1. Performance-approach goals	3.64	1.36	1-7	3.80	1.44	1-6.5	3.63	1.57	1-7	—	.18*	.06
2. Performance-avoidance goals	5.57	1.04	2.5-7	5.90	0.87	4-7	4.95	1.30	2-7	—	—	-.04
3. Self-efficacy for exercises	5.00	1.10	3-7	4.80	1.27	1-7	4.45	1.22	1-6	—	—	—

* $p < .05$.

Table 4
Experiment 2: Multilevel Model Testing the Impact of Graded Versus Nongraded Assessment on Students' Pretask Performance-Avoidance Goals, Controlling for Effects of Student Class

Variable	Model 1		Model 2	
	Estimate	SE	Estimate	SE
Fixed effects ^a				
Intercept	5.42*** ^b	0.11	5.45***	0.09
Grade contrast			-0.28***	0.11
Comment contrast			-0.17	0.11
Self-efficacy (task)			-0.08	0.07
Grade Contrast × Self-Efficacy			-0.08	0.05
Comment Contrast × Self-Efficacy			0.03	0.09
Random effects ^c				
Intercept	0.10 ^d	0.08	0.00	0.00
Residual	1.19*** ^e	0.13	1.10***	0.14
Model fit statistics ^a				
Deviance	557.29		383.70	
AIC	563.29		399.70	
BIC	572.90		422.71	
ICC	.06			

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; ICC = intraclass correlation coefficient.

^a Using maximum likelihood estimation. ^b Overall sample level of pretask performance-avoidance goal adoption. ^c Using restricted maximum likelihood estimation. ^d Variance component corresponding to random intercept. ^e Class-level error.

*** $p < .001$.

avoidance than the other two conditions. This effect was not observed on performance-approach goals, which is consistent with Experiment 1.

In addition, providing evidence in favor of Elliot's (1999) theory concerning the potential overriding effect that salient environmental variables might have over dispositional variables in the prediction of goal adoption, pretask perceptions of perceived competence in class exercises did not influence the adoption of performance-avoidance goals. Nor did they moderate or suppress the relationship between assessment style and pretask performance-avoidance goals, a result that would certainly seem to imply that graded assessments trigger higher levels of pretask performance-avoidance goals in the class as a whole, rather than uniquely among the habitual low achievers of the class. Perceived competence in class exercises, however, measures students' perceptions of their competence in class exercises in general, which in a language class include reading and listening comprehension, use of English, writing, and speaking exercises. As such, it could be argued to constitute an intermediate-level measure (Bandura, 1997), which may not provide sufficient precision (Pajares, 1996), considering that the task in question was one specific language task. Thus, in Experiment 3, we aimed at replicating the results obtained in Experiment 2 with the inclusion of a measure of the students' perceived competence in the specific exercise being assessed.

More importantly, Experiment 3 was devised to test our second general hypothesis, that the relationship between anticipating receiving a grade and pretask performance-avoidance goals will be mediated by the lower levels of autonomous motivation experienced in graded assessment compared with conditions in which no graded assessment was present.

Experiment 3

To test this hypothesis, we set up a longitudinal experiment in which we experimentally manipulated an experience of a graded versus nongraded exercise, then measured autonomous motivation for the task, and finally measured subsequent pretask performance-avoidance goals in a similar task.

Method

Participants. Ninety-seven students attending a Swiss professional school, with a mean age of 18.82 years ($SD = 1.84$; range: 16–24), participated in this experiment. The sample consisted of 38 male and 45 female students (and 14 missing values) in six classes, with two classes allocated to each of the three conditions: grade only, comment only, and grade with comment. This gave 31 students in the grade-only condition, 33 in the comment-only condition, and 33 in the grade-with-comment condition.

Procedure. The first stage of the procedure of Experiment 3 consisted of the students filling out a preliminary questionnaire on motivation for the subject (Midgley et al., 1997), EFL in this case. It was carried out in class in the week preceding the first exercise. This was to be used as a control measure for general subject-level motivation and was included because in French-speaking Swiss professional schools, EFL is not one of the three core subjects (viz. French, maths, and German). At the same time, students filled out a questionnaire on their perceived self-efficacy for the specific tasks they were to carry out in this EFL class, namely reading comprehension and use of English. After they had filled out this questionnaire, the teacher announced the assessment task and assessment conditions for the following week.

The next stage was the same as that used in Experiment 2, with the experiment once again being carried out during a regular EFL class and presented to the students with the collaboration of the class teacher. The previous week, after the students had filled in the baseline measures of motivation for the subject and task-specific self-efficacy, the teacher had informed the students that this week they would carry out a reading comprehension and related use of English exercise to be handed in and assessed with either a grade, individual comment, or grade with comment. At the beginning of this class, the teacher reminded the students of this; the students then filled in the achievement goal questionnaire, and the class carried out the class activity, a reading text with five exercises to test understanding of the text and related use of English skills. During this time, the researcher was present in the classroom. As students finished the activity (all finished well within the 90-min period), they filled in the questionnaire for autonomous motivation (Reported Autonomy Index; Grolnick & Ryan, 1987), translated into French and adapted to task level, about the exercise they had just completed. The researcher collected the questionnaires, and the teacher collected the students'

exercises. The teacher announced that the students would get their assessed exercises back at the beginning of the next class the following week and that they would then carry out a second similar exercise under the same assessment conditions, that is, for a grade, for a grade and comments, or for comment-based feedback. The teacher and researcher, in collaboration, corrected the exercises according to the normal assessment standards used by the teacher.

Corrections in the grade-only condition consisted of ticks for correct answers, crosses for incorrect answers, a score for each section, and an overall grade marked on the front page of the stapled exercise sheets. This procedure in fact was consistent with the assessment pattern normally followed in this subject area. Corrections in the grade-with-comment condition consisted of the same corrections as in the grade-only condition, but in addition, comments were added next to wrong answers that suggested strategies for finding the correct answer (e.g., "Look for 'have to' in the text") or rephrased the question to help the student to rethink the answer (e.g., "At what time do they arrange to meet?" to clarify the question "When do Brad and Kim arrange to meet on Saturday?"). A maximum of two comments was provided per exercise. At the end of the exercise sheet, an overall comment was provided, which pinpointed an area in which the student had mastered the requirements of the task (e.g., "Comprehension of the diary mastered" or "Some original answers") and provided advice on how to further improve reading comprehension skills (e.g., "Always go back to the text for the text comprehension questions and find the key words of the question in it. The answer to the question will often be next to these"). Corrections in the comment-only condition consisted of the same comment-based strategy as in the grade-with-comment condition but with no accompanying ticks, crosses, scores, or final grade.

The following week, in the next class, the teacher returned the corrected exercises. Once the students had had time to look through the corrected exercises, the teacher announced the second exercise, the students filled in a second pretask achievement goal questionnaire, and the students then proceeded to carry out the exercise. At the end of the exercise, students were debriefed, and those who had received grades for the previous week's work were informed that the grade did not have to count for their semester GPA, but could if they wished. Consequently, students were given

a choice as to whether to participate in the questionnaire filling-in and whether to keep the grade allocated for the exercise. Their only obligation was to carry out the class activity. It was agreed with the teacher that for reasons of academic discretion, task performance scores would not be collected for the research.

Measures. Measures of participants' motivation for the subject, pre-Task 2 performance-avoidance goals, posttask autonomous motivation for the exercise, perceived competence in reading comprehension and use of English, and baseline competence in the subject were obtained.

Motivation for the subject. The students' motivation for the subject was measured with nine items from a subject-level version of the Intrinsic Motivation Inventory (McAuley, Duncan, & Tammen, 1989; Self-Determination Theory, 2008; e.g., "I find this class interesting"; $\alpha = .86$). Items were translated into French and then back-translated blind in order to ensure compatibility in meaning between languages. Descriptive statistics are presented in Table 5.

Achievement goals. The students' pre-Task 2 performance-avoidance goals were measured with the same scales as in Experiments 1 and 2. The three-item, pre-Task 2 performance-avoidance goal alpha was satisfactory ($\alpha = .79$) but improved ($\alpha = .93$) when reduced to two items as in the pre-Task 1 scale and in Experiments 1 and 2. Consequently, to maintain consistency with Experiments 1 and 2, we retained the two-item scale.

Autonomous motivation (reported autonomy index). The posttask measure of autonomous motivation, the Reported Autonomy Index (Grolnick & Ryan, 1987), was derived from the four scales of the Perceived Locus of Causality Questionnaire (Ryan & Connell, 1989), translated and adapted to task level. It included extrinsic motivation with external regulation (e.g., "I did the exercise mainly because I had to"; $\alpha = .87$), extrinsic motivation with introjected regulation (e.g., "I did the exercise mainly because I want the teacher to think I'm a hardworking student"; $\alpha = .85$), extrinsic motivation with identified regulation (e.g., "I did the exercise mainly because I think it will be useful for me"; $\alpha = .82$), and intrinsic motivation (e.g., "I did the exercise mainly because it's interesting"; $\alpha = .92$). These four scales were weighted according to their degree of autonomous motivation (external regulation, -2 ; introjected regulation, -1 ; identified regulation, 1 ;

Table 5
Experiment 3 ($N = 97$): Descriptive Statistics and Zero-Order Correlations Among Variables

Variable	Grade only			Grade and comment			Comment only			1	2	3	4	5	6
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range						
1. Task 1 autonomous motivation	-1.06	4.98	-9.9-9	-1.87	4.36	-9-9	0.75	3.79	-7-10.4	—	-.18	-.37**	.29*	.39***	.33**
2. Pre-Task 2 performance-approach goals	3.19	1.61	1-6	2.58	1.88	1-7	2.35	1.53	1-6.5	—	.21	.30*	.33**	.21	
3. Pre-Task 2 performance-avoidance goals	5.58	1.22	2-7	5.43	1.32	2.5-7	4.06	1.70	1-7	—	-.14	.27*	-.11		
4. Self-efficacy for exercises	4.90	1.28	2-7	5.45	1.10	3-7	4.94	1.60	1-7	—	.40***	.74***			
5. Subject-level intrinsic motivation	5.18	1.39	2.5-7	5.21	0.95	3.5-7	4.57	1.14	2-7	—				.37**	
6. Last semester subject grade point average	4.32	1.08	2-6	4.95	1.09	2.5-6	4.31	1.45	2-6	—					

* $p < .05$. ** $p < .01$. *** $p < .001$.

intrinsic motivation, 2; Grolnick & Ryan, 1987), the weights being justified by the simplex-like structure of the four scales (Guttman, 1954, cited by Grolnick & Ryan, 1987). The final score was calculated by adding the weighted scores together to create a continuous variable reflecting less to more autonomous motivation (Grolnick & Ryan, 1987). Items in the scales were translated into French and then back-translated blind in order to ensure compatibility in meaning between languages.

Perceived competence in reading comprehension and use of English (task-specific self-efficacy). Students' perceptions of competence in reading comprehension and use of English were measured with a two-item scale ("I can do the reading comprehension exercises in this class well," "I can do the use of English exercises in this class well"; $\alpha = .88$). Although task-specific self-efficacy was measured along with motivation for the subject the week before the first assessment and prior to the teacher's forewarning of the up-and-coming activity and assessment conditions and, as such, could not have been affected by the experimental condition, to control for random differences between conditions, we ran an analysis of variance testing the impact of assessment condition on task-level self-efficacy. Results of the analysis were not significant, $F(2, 130) = 2.51, ns$.

Baseline competence. An external measure of students' baseline competence was obtained via their subject GPA scores provided by the school administration and was used as a statistical control in supplementary analyses. The score ranged from 1 to 6, with 4 as the pass level.

Results

Overview of the analyses. Two sets of analyses were carried out to test our two hypotheses. Once again we used a mixed-effects model in SPSS, as experimental conditions were made up of classes. The first set of analyses tested the impact of assessment conditions on pre-Task 2 performance-avoidance and performance-approach goals. To test the effects of the grade and comment-based teacher feedback, we created the same two contrasts as in Experiment 2, using Helmert coding: Contrast 1, the grade contrast, with the comment-only feedback condition coded 2 and the grade-with-comment and grade-only conditions coded -1 ; and Contrast 2, the comment contrast, which tested the impact of grade-with-comment versus grade alone, with the comment-only condition coded 0, the grade-with-comment condition coded -1 , and the grade-only condition coded 1. Once again we tested the potential moderating impact of self-efficacy, this time not only task level but also task specific. Moreover, student perceptions of intrinsic motivation for the subject (EFL) were included in the fixed effects, as preliminary analyses revealed a significant difference in intrinsic motivation for subject between the two conditions tested in the grade contrast, $B = -0.21, F(1, 94) = 5.88, p < .05, \eta^2 = .06$. Students in the grade conditions revealed higher levels of subject-level intrinsic motivation ($M = 5.20, SD = 1.18$) than those in the no-grade condition ($M = 4.57, SD = 1.13$).

The second set of analyses consisted of mediation analyses carried out to test Hypothesis 2 and see whether the degree of autonomous motivation experienced in Task 1 explained potential differences between assessment conditions in pre-Task 2 performance-avoidance goal adoption. Results are reported in Table 6. Further to this, the

same analyses were run with, this time, student subject-level GPA acting as a potential moderator of the assessment condition to pretask performance-avoidance goal relationship. These results are reported in Table 7. As in Experiments 1 and 2, we divided the first set of analyses into two stages, the first focusing on variance explained by class as opposed to fixed effects and the second focusing on the relative impact of the fixed effects and comparing model fit.

Hypothesis 1: Stage 1: Class variance. Results of our first unconditional model, with class entered as a random effect, indicated a nonsignificant contribution to performance-avoidance goal adoption by class (Wald $Z = 1.20, ns$) and between-class residual ($\sigma^2 = 1.96, SE = 0.34, p < .001$) over 3 times that of the between-class variance ($\tau^2 = 0.59, SE = 0.47$), implying that the majority of variance was unexplained in this model. The ICC indicated that when the experimental conditions in which classes were nested were not included in the model, approximately 23% of the variance in performance-avoidance goal adoption was explained by the class.

In Model 2, fixed effects were added, including the grade contrast and the comment contrast, in order to test the impact of the two contrasts on pretask performance-avoidance goals. Task-specific self-efficacy was also included as a potential moderator of the impact of each contrast on pretask performance-avoidance goal adoption, and thus interactions between task-specific self-efficacy and the two contrasts were also added to the model. Finally, student perceptions of intrinsic motivation for the subject (EFL) were included in the fixed effects. This gave six fixed effects. Class was included as a random effect to control for the contribution of class to pretask performance-avoidance goal adoption. Results of Model 2 indicate a large reduction in between-class variance ($\tau^2 = 0.00, SE = 0.00, ns$) and a highly significant between-class residual ($\sigma^2 = 1.67, SE = 0.29, p < .001$).

Hypothesis 1: Stage 2: Fixed effects and model comparison. In Stage 2, the focus was on testing the contribution of our fixed effects on performance-avoidance goal adoption and testing the relative fit of the unconditional Model 1 with the mixed Model 2 to the data. Results of Stage 2 analyses indicate that the grade contrast had a significant effect on pre-Task 2 performance-avoidance goals ($\gamma = -0.33, SE = 0.12, p < .01$), with significantly lower levels in the no-grade condition ($M = 4.06, SD = 1.70$) than in the two grade conditions ($M = 5.49, SD = 1.26$). In addition, intrinsic motivation for the subject was significantly associated with pre-Task 2 performance-avoidance goals ($\gamma = 0.46, SE = 0.17, p < .01$), and there was a main effect of task-specific self-efficacy on pre-Task 2 performance-avoidance goals ($\gamma = -0.39, SE = 0.14, p < .01$). The interaction between task-specific self-efficacy and the grade contrast was not significant, and no other fixed effects reached significance. Deviance in Model 1, the unconditional model, was 257.44. Deviance in Model 2, the model including fixed as well as random effects, dropped to 224.50, a significant difference of 32.94, $\chi^2(6, N = 97) = 32.94, p < .001$, indicating that Model 2 with fixed effects provided a significantly improved fit to the data. In sum, results of our first set of analyses run on pre-Task 2 performance-avoidance goals satisfy the first step of the mediation process (Baron & Kenny, 1986), with the grade contrast significantly influencing pre-Task 2 performance-avoidance goal adoption.

Our first set of analyses was, as in Experiments 1 and 2, also run on pre-Task 2 performance-approach goals. When the same Stage 1 and 2 analyses were run on pre-Task 2 performance-approach

Table 6

Experiment 3: Multilevel Model Testing the Mediation by Task 1 Autonomous Motivation of the Relationship Between Assessment Conditions and Pre-Task 2 Performance-Avoidance Goals, Controlling for Effects of Student Class (Task-Specific Self-Efficacy as Moderator)

Variable	Model 2							
	Model 1: Pre-Task 2 performance-avoidance goals		Pre-Task 2 performance-avoidance goals		Task 1 autonomous motivation		Model 3: Pre-Task 2 performance-avoidance goals	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Fixed effects ^a								
Intercept	-0.00 ^b	0.36	-0.04	0.16	-0.14	0.47	-0.05	0.16
Grade contrast			-0.33**	0.12	1.07*	0.34	-0.19	0.13
Comment contrast			-0.01	0.21	0.40	0.58	0.04	0.20
Self-efficacy (task)			-0.39	0.14	0.45	0.36	-0.29*	0.13
Intrinsic motivation (subject)			0.49**	0.17	1.50***	0.40	0.63***	0.18
Grade Contrast × Self-Efficacy			0.13	0.08	0.03	0.22	0.12	0.08
Comment Contrast × Self-Efficacy			-0.15	0.17	0.41	0.44	-0.12	0.16
Task 1 autonomous motivation							-0.12**	0.04
Random effects ^c								
Intercept	0.59 ^d	0.47	0.00	0.00	0.18	0.74	0.00	0.00
Residual	1.96***	0.34	1.67***	0.29	13.94***	2.31	1.48***	0.27
Model fit statistics ^a								
Deviance	257.44		224.50		433.30		196.83	
AIC	261.44		242.50		451.30		216.83	
BIC	269.93		262.34		472.63		237.94	
ICC	.23							

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; ICC = intraclass correlation coefficient. ^a Using maximum likelihood estimation. ^b Overall sample level of pre-Task 2 performance-avoidance goal adoption (0 as centered). ^c Using restricted maximum likelihood estimation. ^d Variance component corresponding to random intercept. ^e Class-level error. * $p < .05$. ** $p < .01$. *** $p < .001$.

goals, Stage 1 analyses produced an ICC of 5.8%, variance due to class being nonsignificant ($\gamma = 0.14$, $SE = 0.23$, Wald $Z = 0.61$, ns). Among Stage 2 fixed-effect contrasts, the grade contrast had no effect on pretask performance-approach goals ($\gamma = -0.19$, $SE = 0.84$, ns), which remained strictly similar in graded ($M = 2.89$, $SD = 1.75$) versus nongraded conditions ($M = 2.35$, $SD = 1.53$). No other significant effects appeared.

Hypothesis 2: Mediation analyses, Model 2. To test the second step of the mediation process, we reran Model 2 using uniquely MLE with Task 1 autonomous motivation as dependent variable. Results revealed that the grade contrast had a significant effect on Task 1 autonomous motivation ($\gamma = 1.07$, $SE = 0.34$, $p < .05$), with higher levels of Task 1 autonomous motivation being experienced in the no-grade condition ($M = 0.75$, $SD = 3.79$) compared with the grade conditions ($M = -1.16$, $SD = 4.64$). Intrinsic motivation for the subject predicted Task 1 autonomy ($\gamma = 1.50$, $SE = 0.40$, $p < .001$). No other significant effects appeared, and random effects, run in a preliminary unconditional model with RMLE, revealed that variance in Task 1 autonomous motivation due to class was nonsignificant ($\tau^2 = 0.60$, $SE = 1.39$, Wald $Z = 0.43$, ns), with an intraclass correlation of 2.9%.

Hypothesis 2: Mediation analyses, Model 3. Finally, to test the third step of the mediation process, we ran Model 3 on pre-Task 2 performance-avoidance goals. Model 3 replicated the first step of

Model 2 in all aspects except for the inclusion of the hypothesized mediating variable, Task 1 autonomous motivation. This enabled us to test for the potential mediating role of Task 1 autonomous motivation on the grade contrast to pre-Task 2 performance-avoidance goal relationship. Model 3 consequently contained seven fixed effects. Random effects of class were the same as in Model 2.

Results indicated that Task 1 autonomous motivation had a significant impact on pre-Task 2 performance-avoidance goals ($\gamma = -0.12$, $SE = 0.04$, $p < .01$). Concurrent with this, the impact of the grade contrast on pre-Task 2 performance-avoidance goals was no longer significant ($\gamma = -0.19$, $SE = 0.13$, ns), indicating that Task 1 autonomous motivation had mediated its effect on pre-Task 2 performance-avoidance goals. A Sobel test ($Z = -2.17$, $SE = 0.06$, $p < .05$) confirmed that the mediation effect was indeed significant. Results are presented graphically in Figure 1. Task-specific self-efficacy and intrinsic motivation for the subject both predicted pre-Task 2 performance-avoidance goals in this model ($\gamma = -0.29$, $SE = 0.13$, $p < .01$, and $\gamma = 0.63$, $SE = 0.18$, $p < .001$, respectively). No other significant effects emerged.

Random effects also did not change, with variance in pre-Task 2 performance-avoidance goals due to class remaining nonsignificant ($\tau^2 = 0.00$, $SE = 0.00$, ns). Deviance in Model 3 dropped from 224.50 to 196.83, a significant difference of 26.67, $\chi^2(1, N = 97) = 26.67$, $p < .001$, indicating that Model 3 with the addition

Table 7

Experiment 3: Multilevel Model Testing the Mediation by Task 1 Autonomous Motivation of the Relationship Between Assessment Conditions and Pre-Task 2 Performance-Avoidance Goals, Controlling for Effects of Student Class (Subject Grade Point Average as Moderator)

Variable	Model 1: Pre-Task 2 performance-avoidance goals		Model 2				Model 3: Pre-Task 2 performance-avoidance goals	
			Pre-Task 2 performance-avoidance goals		Task 1 autonomous motivation			
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Fixed effects ^a								
Intercept	-0.00 ^b	0.36	-0.09	0.15	0.36	0.51	-0.05	0.21
Grade Contrast			-0.31*	0.14	1.49***	0.41	0.18	0.18
Comment contrast			-0.10	0.19	0.52	0.41	0.05	0.22
Subject grade point average			-0.61***	0.16	1.10*	0.46	-0.46*	0.17
Intrinsic motivation (subject)			0.59***	0.13	1.48***	0.43	0.72***	0.17
Grade Contrast × Grade Point Average			0.14	0.10	0.30	0.32	0.15	0.10
Comment Contrast × Grade Point Average			-0.24	0.17	0.72	0.51	-0.14	0.17
Task 1 autonomous motivation							-0.11**	0.04
Random effects ^c								
Intercept	0.59 ^d	0.47	0.00	0.00	0.00	0.00	0.06	0.11
Residual	1.96*** ^e	0.34	1.29***	0.25	13.90***	2.44	1.13***	0.24
Model fit statistics ^a								
Deviance	257.44		170.27		355.54		153.03	
AIC	261.44		188.27		373.54		173.03	
BIC	269.93		206.33		393.11		192.35	
ICC	.23							

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; ICC = intraclass correlation coefficient. ^a Using maximum likelihood estimation. ^b Overall sample level of pre-Task 2 performance-avoidance goal adoption (0 as centered). ^c Using restricted maximum likelihood estimation. ^d Variance component corresponding to random intercept. ^e Class-level error. * $p < .05$. ** $p < .01$. *** $p < .001$.

of Task 1 autonomous motivation provided a significantly improved fit to the data.

The same mediation testing procedure was run on pre-Task 2 performance-avoidance goals, this time including subject GPA as a potential moderating variable. Results, reported in Table 7, reveal the same effects as those obtained with task-specific self-efficacy as a moderating variable. A Sobel test ($Z = -2.19, SE =$

$0.08, p < .05$) confirmed that the mediation effect was once again significant.

As a final comment, it should be noted that pre-Task 1 measures of performance-avoidance goals could not be included as a baseline measure of performance-avoidance orientation, as they were taken after the experimental conditions had been announced and were specific to the anticipated task type. However, as a supplementary analysis carried out in order to control for relations between pre-Task 1 and pre-Task 2 performance-avoidance goals, we did test a three-level multilevel model with Level 1 being the intrapersonal measure of performance-avoidance goals pre-Task 1 and pre-Task 2. Results of this replicated the results presented in the two-level model above.¹

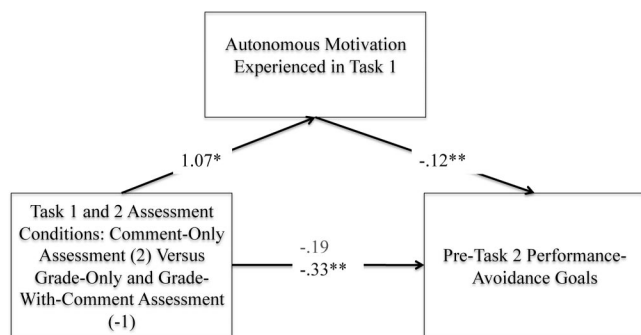


Figure 1. Experiment 3: Mediation by Task 1 autonomous motivation of the relationship between assessment conditions and pre-Task 2 performance-avoidance goals. All values represent unstandardized coefficients. * $p < .05$. ** $p < .01$.

¹ In order to control for relations between pre-Task 1 and pre-Task 2 performance-avoidance goals, we created a three-level multilevel model with Level 1 being the intrapersonal measure of performance-avoidance goals pre-Task 1 and pre-Task 2. Treating the wave variable (performance-avoidance goal adoption at Time 1 and Time 2) as a categorical variable allowed us to pinpoint the impact of the grade and comment contrasts on pre-Task 1 and pre-Task 2 performance-avoidance goals separately, while at the same time controlling for correlations between the two sets of performance-avoidance goals via the autoregressive covariance matrix. Level 2 dealt with variation arising from the fixed effects, experimental contrasts, motivation for subject, and self-efficacy, and Level 3 dealt with random class effects.

Discussion

Results of Experiment 3 again revealed the effect of graded versus nongraded assessment on pre-Task 1 performance-avoidance goals, after controlling this time for potential moderation effects by task-specific self-efficacy and GPA. This effect conceptually replicates that of Experiment 1, and directly replicates that of Experiment 2, with a longitudinal design and by controlling for the moderation effect of a measure of perceived competence specifically mentioning the exercises at hand. Importantly, the longitudinal nature of Experiment 3 allows one to see that the reduction in autonomous motivation experienced by students in the grade conditions in Task 1 played a mediational role in the relationship between assessment conditions and pre-Task 2 performance-avoidance goals. That is, the experience of reduced autonomous motivation in graded assessment compared with nongraded assessment conditions can help explain why students adopt pretask performance-avoidance goals when faced with graded tasks.

Results of our first unconditional model, with class entered as a random effect and repeated measures of pre-Task 1 and 2 entered as a repeated effect, indicated a nonsignificant contribution to performance-avoidance goal adoption by class ($\tau^2 = 0.58$, $SE = 0.42$, Wald $Z = 1.38$, *ns*). Repeated measures of pre-Task 1 and 2 revealed that the two measures were highly correlated ($\tau^2 = 0.53$, $SE = 0.09$, Wald $Z = 5.79$, $p < .001$).

In Model 2, repeated measures of pretask performance goals were once again included, and fixed effects were added. These consisted firstly of the variable wave that was entered to provide the intercept for pre-Task 1 (Time 1) and pre-Task 2 (Time 2) performance-avoidance goals. We then included the grade contrast and comment contrasts in interaction with wave to test their impact on pre-Task 1 and 2 performance-avoidance goals separately. Task-specific self-efficacy was also included as a potential moderator of the impact of each contrast on pretask performance-avoidance goal adoption, and interactions between task-level self-efficacy and the two contrasts were also added to the model. Finally, student perceptions of intrinsic motivation for the subject (EFL) were included in the fixed effects, as in the main analysis. This gave eight fixed effects. Class was included as a random effect to control for the contribution of class to pretask performance-avoidance goal adoption.

Results indicate that the grade contrast had a significant effect on pre-Task 1 performance-avoidance goals ($\gamma = -0.34$, $SE = 0.12$, $p < .05$), with significantly lower levels in the no-grade condition ($M = 4.45$, $SD = 1.33$) than in the two grade conditions ($M = 5.69$, $SD = 1.18$). The grade contrast also had a significant effect on pre-Task 2 performance-avoidance goals ($\gamma = -0.39$, $SE = 0.12$, $p < .01$), with significantly lower levels in the no-grade condition ($M = 4.06$, $SD = 1.70$) than in the two grade conditions ($M = 5.49$, $SD = 1.26$). In addition, intrinsic motivation for the subject was significantly associated with pretask performance-avoidance goals ($\gamma = 0.24$, $SE = 0.12$, $p < .05$). No other fixed effects reached significance.

Task 1 autonomy was then added to the model in interaction with wave to specify its impact on pre-Task 1 and pre-Task 2 performance-avoidance goals separately. It had a significant impact on pre-Task 2 performance-avoidance goals ($\gamma = -0.12$, $p < .001$). Concurrent with this, the impact of the grade contrast on pre-Task 2 performance-avoidance goals was reduced ($\gamma = -0.29$, $p < .05$), indicating that Task 1 autonomy had mediated its effect on pre-Task 2 performance-avoidance goals. A Sobel test ($Z = -2.42$, $SE = 0.06$, $p < .05$) confirmed that the partial mediation effect was indeed significant. As expected, Task 1 autonomy had no effect on pre-Task 1 performance-avoidance goals ($\gamma = -0.05$, $SE = 0.03$, *ns*).

It should be noted that both task-specific self-efficacy and subject GPA had a negative effect on pretask performance-avoidance goal adoption. The main effect of task-specific self-efficacy reflects Elliot and Moller's (2003) argument that perceived competence may act as a general antecedent of performance-avoidance goal adoption, as well as work by Cury et al. (2006) that showed that perceived competence negatively predicts general performance-avoidance goal adoption rather than plays a moderating role in the relationship between achievement goals and other variables. The main effect of subject GPA echoes findings that higher academic standing, as measured by general GPA, has a negative impact on generalized performance-avoidance goal adoption (Hsieh, Sullivan, & Guerra, 2007). This is not surprising, as higher academic standing constitutes a generalized buffer when it comes to performance concerns. What is, of course, much more interesting is that, despite this, all levels of task-specific self-efficacy and subject-level GPA are more apt to adopt performance-avoidance goals in graded versus nongraded assessment contexts.

General Discussion

Three experiments were carried out to test two main hypotheses. First, we hypothesized that conditions in which graded feedback is anticipated compared with conditions in which graded feedback is not anticipated will induce higher levels of pretask performance-avoidance goals, a difference that should not appear for pretask performance-approach goals. Our second hypothesis was that the relation between anticipating graded or nongraded teacher feedback and pretask performance-avoidance goals will be mediated by the lower levels of autonomous motivation previously experienced in conditions characterized by graded assessment compared with conditions in which no graded assessment was present. These hypotheses address an important question in view of, on the one hand, the prevalence of testing throughout the educational system (Ryan & Weinstein, 2009) and, on the other, the resolutely negative effect of performance-avoidance goals on behavioral and performance-related outcomes, an effect that has been described by McGregor and Elliot (2002, p. 393) as "inimical throughout the exam sequence" and potentially "devastating."

Results reveal that, as predicted, anticipating graded feedback as opposed to no feedback (Experiment 1) or comment-based feedback (Experiments 2 and 3) increased adoption of pretask performance-avoidance goals while having no significant effect on performance-approach goal adoption. The replication of the effect of increased pretask performance-avoidance goal endorsement in graded versus nongraded conditions throughout the three experiments, and its persistence in the more stringent comparison of graded versus nongraded teacher assessment applied in Experiments 2 and 3, attest to the robustness of the phenomenon.

Although such a result may seem surprising in the light of the common assumption that a function of grading is to stimulate competition and hence inspire people to give of their best (see Elliot & Moller, 2003; Romanowski, 2004), a close reading of salient literature provides a number of indications that provide substantial support for the grade to performance-avoidance goal relationship. According to Harter's (1978) interpretation of post hoc student comments, having class work graded increased worry over the possibility of getting a bad grade, an argument echoed by

Ames and Ames (1981), who claimed that normative evaluation criteria reduce the chances of developing a positive achievement orientation, and also by Butler and Nisan (1986), according to whom grades engender a fear of failure. However, despite these indications, to our knowledge, the findings presented in this research constitute the first direct, experimental test of the link between grades and pretask performance-avoidance goals.

Ancillary to the grade to performance-avoidance goal relationship result is our second finding of interest: that no significant differences in pretask performance-avoidance goal adoption emerged between the grade-only and the grade-with-comment conditions (Experiments 2 and 3). Compared with the comment-only condition, both provoked significantly higher levels of performance-avoidance goals and lower levels of autonomous motivation. This result is of theoretical interest in that, 20 years on, it extends Butler's (1988) parallel finding that grades and grades accompanied by comments incited equally lower levels of intrinsic motivation. It would seem that when comment-based assessment is accompanied by a grade, the potentially beneficial impact of the comment on pretask performance-avoidance goals is overshadowed by the grade.

This result is equally noteworthy from a practical standpoint when one considers the widespread support in educational literature for comment-based feedback (Black & William, 1998; Higgins, Hartley, & Skelton, 2002), which, combined with pressure from educational authorities to give grades (Nichols & Berliner, 2008), is all too likely to push conscientious and caring educators to spend time adding comments alongside the grade they have attributed to a piece of work. However, results of this study and of Butler's (1988) imply that simply adding a comment to a graded piece of work may not provide the motivational benefits educators may hope for, although, of course, performance-related benefits may indeed ensue from the provision of a comment with a grade. Further research would do well to clarify this point.

A third finding resulting from our studies was that the lower levels of autonomous motivation experienced in the conditions characterized by graded assessment effectively mediated the grade to pretask performance-avoidance goal relationship in an ulterior task of the same nature (Experiment 3). This finding constitutes a significant contribution to the literature in two ways. On the one hand, from a methodological standpoint, the mediation of the grade-goal relationship by autonomous motivation provides support for the argument that autonomous motivation acts as a psychological mediator between environmental factors and states of motivation (Grouzet, Vallerand, Thill, & Provencher, 2004). This is compatible with achievement goal theory, which also argues in favor of the existence of mediators that explain the link between environmental variables and achievement goal endorsement (Elliot, 1999). In this way, results of Experiment 3 furnish a timely integration of achievement goal and self-determination theory and, furthermore, support for the multitiered motivational models favored by a number of theorists (DeShon & Gillespie, 2005; Elliot, 1999; Grouzet et al., 2004; Moller & Elliot, 2006).

On the other hand, from a more theoretical perspective, the mediation of the grade-goal relationship by autonomous motivation provides a convincing explanation of the negative impact of grading on pretask performance goal adoption. A long-standing tradition of research has associated grading with a lack of autonomous motivation (Butler & Nisan, 1986; Deci & Ryan, 1985;

Grolnick & Ryan, 1987; Harter, 1978; Ryan, 1982), via its role in creating pressure to do and pressure to achieve, in many instances competitively. According to Covington (2000, p. 181), grades constitute "the unmistakable measure by which many, if not most youngsters judge their worth as students," a direct consequence no doubt of "society's tendency to equate worth with the ability to achieve competitively" (Covington & Omelich, 1979, p. 688). Our findings take these arguments a stage further by establishing a potential link between the powerlessness and dependence caused by external grading that is interpreted at the individual level as controlled motivation and the consequent focus on avoiding negative outcomes. We would argue that it is not simply competition but competition allied with powerlessness that swings the focus of attention over to performance-avoidance goals prior to engaging in a graded task.

The fourth finding of interest in this research was that although, in Experiment 3, task-specific perceived competence did indeed reduce pretask performance-avoidance goal adoption in general, it did not play a significant moderation role in the grade to pretask performance-goal relationship. In this respect, the results of Experiments 2 and 3 provide further support for Elliot's (1999) argument that the association between low competence and performance-avoidance goal adoption is likely to be overridden by salient environmental cues. In line with this, neither of our two measures of perceived competence, in Experiments 2 and 3, nor our measure of externally attributed competence (subject GPA), in Experiment 3, had any moderating impact on the grade-pretask goal relationship.

Why might the negative consequences of grading on pretask goals not be reserved uniquely for chronic low achievers? Grading, particularly when normative, has been argued to generate high levels of uncertainty (Covington & Omelich, 1984) over the criteria for success and consequently over the potential success or failure of future performance. Furthermore, the potentially threatening nature of this uncertainty in the face of a graded evaluation is doubly heightened, as, unlike in the case of certain competitive activities, the stakes are not simply zero sum, as Covington (2000) claimed, but potentially "minus sum"—underperforming will leave you worse off than before. This is because most graded evaluations carry considerable ongoing consequences: Either they are part of a continuous assessment program, such as the GPA system, or else they have a direct impact on academic progress, as in the case of standardized tests. As a result of this, directly experienced pressure to do and to succeed in such a high-stakes context as a graded evaluation, which Covington (2000, p. 185) referred to as a "failure-oriented (competitive) ability game," has every chance of seriously increasing failure-focused attention (Covington, 2000) even for high-achieving students, which could in turn lead to choking under pressure (Beilock & Carr, 2001). That is, the salience of potential failure that orients achievement contexts toward threat and can consequently generate performance-avoidance goals (Elliot, 1999) is a relevant issue in the context of graded assessment for all students, whatever their ability range.

One limitation of this research is the measure of performance-avoidance used. As discussed earlier, since the experiments were carried out, Elliot and Murayama (2008) have rehabilitated the original scale, eliminating the item that did not in our studies, as well as in other studies (Darnon et al., 2007), cohere with the other items. For this reason and the fact that, in Experiment 3, the

two-item scale was indeed in line with preexisting theory and results, predicted by task-level self-efficacy and subject GPA, we would argue that although a two- as opposed to three-item scale is a limitation on comparative potential with other studies, the results provided by the two-item scale are relevant to the field. Of course, future studies should aim to reexplore this area using the updated scale.

A second limitation is the lack of a control group to compare the comment group to, which precludes the possibility of drawing conclusions about the positive impact of comment-based feedback. However, as feedback on work is an essential part of the learning process (Butler, 1987, 1988), and most feedback by default has to be either in the form of a symbolic value such as a grade or in the form of individualized comments, we argue for the value of the comment when unaccompanied by a grade. Furthermore, this study focused uniquely on the impact of assessment style on pretask performance goals and autonomous motivation. Future research would do well to focus on establishing what effects on these variables comment-based feedback provides when compared with no feedback. In addition, it is worth exploring the longitudinal impact of the individualized comment, written or spoken, on other measures of motivation, as well as task performance.

A third point worth mentioning is that the grade conditions in these experiments were linked to an ongoing subject GPA and, consequently, more was at stake than would be in a one-off graded condition in a laboratory setting. However, we would argue that the whole significance and power of the graded evaluation is the weight of consequences that it carries for the individual, and as such, the ecological nature of these experiments is a strength. Nevertheless, it would be interesting to retest this phenomenon in the purer air of a laboratory setting to see whether the effect still held.

Finally, it is important to bear in mind the limited scope of the state achievement goals measured in this research. We have focused on pretask performance goals, but this is only one short, though important, stage of the classroom assessment process (Brookhart & DeVoge, 1999). It is likely that other goals, particularly mastery-approach goals, may indeed be more salient at other points in the cycle, and interact with other goals (Barron & Harackiewicz, 2001; Darnon, Dompnier, Gilliéron, & Butera, 2010). In relation to this, our research is a first step in the exploration of the longitudinal impact of grades on motivation and goals and, as such, works with a specific age segment, subject area, and educational context. Consequently, further research that replicates and extends the scope of these studies would help enormously in taking our understanding of this area still further.

In addition to the research possibilities outlined above, the results of Experiment 3, which indicate that perceived task autonomy mediates the relation between assessment style and subsequent pretask performance-avoidance goals, point the way to a potentially fruitful, new avenue of research. These results, when taken in conjunction with those of McGregor and Elliot (2002), which show that pretask performance-avoidance goal adoption predicts increased test anxiety, desire to escape, reduced ability-related self-esteem, and competence expectancies, are consistent with those of Harter's (1978) self-determination-related explanation of why grades make children worry in a task accomplishment situation. They are also consistent with more recent research findings of Dickerson and Kemeny (2004), whose meta-analysis

revealed that socioevaluative threat, engendered by the presence of an evaluative audience in the context of a motivated performance situation in which core attributes such as competence and intelligence are at stake, when combined with conditions perceived to be uncontrollable, produced the highest bodily levels of cortisol in stress response hormonal activation.

Taken as a whole, our results and these could imply that perceived task autonomy really could be playing a key role in the grade to preperformance to avoidance goal relationship. If this is the case, then we have a golden opportunity for exploring ways in which to attenuate this association. It would be indeed an important and useful undertaking to find out, for example, whether it is the essentially summative nature of a grade (Bloom, Hasting, & Madaus, 1971) or the increased diagnosticity (Elliot & Moller, 2003) and resultant labeling (Thomas & Oldfather, 1997) or the social accountability commonly associated with grades, or indeed any combination of these, that drives perceptions of reduced autonomy in the context of a graded task and subsequent pretask performance-avoidance goal adoption. Assessments and grades occupy a central position in students' lives (Pope, 2001). Consequently, seeking ways to make them as user-friendly as possible is a most worthy quest for researchers interested in education.

References

- Albright, J. J., & Marinova, D. M. (2010). *Estimating multilevel models using SPSS, Stata, SAS, and R*. Retrieved from <http://www.iub.edu/~statmath/stat/all/hlm/hlm.pdf>
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261–271. doi:10.1037/0022-0663.84.3.261
- Ames, C., & Ames, R. (1981). Competitive versus individualistic goal structures: The salience of past performance information for causal attributions and affect. *Journal of Educational Psychology, 73*, 411–418. doi:10.1037/0022-0663.73.3.411
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173
- Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology, 80*, 706–722. doi:10.1037/0022-3514.80.5.706
- Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130*, 701–725. doi:10.1037/0096-3445.130.4.701
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*, 7–74. doi:10.1080/0969595980050102
- Bloom, B. S., Hasting, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.
- Boggiano, A. K., & Barrett, M. (1985). Performance and motivational deficits of helplessness: The role of motivational orientations. *Journal of Personality and Social Psychology, 49*, 1753–1761. doi:10.1037/0022-3514.49.6.1753
- Bressoux, P. (2008). *Modélisation statistique appliquée aux sciences sociales* [Statistical modeling applied to social sciences]. Brussels, Belgium: De Boeck.
- Brookhart, S. M., & DeVoge, J. G. (1999). Testing a theory about the

- role of classroom assessment in student motivation and achievement. *Applied Measurement in Education*, *12*, 409–425. doi:10.1207/S15324818AME1204_5
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology*, *79*, 474–482. doi:10.1037/0022-0663.79.4.474
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*, 1–14.
- Butler, R. (2006). Are mastery goals and ability goals both adaptive? Evaluation, initial goal construction and the quality of task engagement. *British Journal of Educational Psychology*, *76*, 595–611. doi:10.1348/000709905X52319
- Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, *78*, 210–216. doi:10.1037/0022-0663.78.3.210
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171–200. doi:10.1146/annurev.psych.51.1.171
- Covington, M. V., & Müeller, K. J. (2001). Intrinsic versus extrinsic motivation: An approach/avoidance reformulation. *Educational Psychology Review*, *13*, 157–176. doi:10.1023/A:1009009219144
- Covington, M. V., & Omelich, C. L. (1979). It's best to be able and virtuous too: Student and teacher evaluative responses to successful effort. *Journal of Educational Psychology*, *71*, 688–700. doi:10.1037/0022-0663.71.5.688
- Covington, M. V., & Omelich, C. L. (1984). Task-oriented versus competitive learning structures: Motivational and performance consequences. *Journal of Educational Psychology*, *76*, 1038–1050. doi:10.1037/0022-0663.76.6.1038
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*, 438–481. doi:10.3102/00346543058004438
- Cury, F., Elliot, A. J., Da Fonseca, D., & Moller, A. C. (2006). The social-cognitive model of achievement motivation and the 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *90*, 666–679. doi:10.1037/0022-3514.90.4.666
- Darnon, C., & Butera, F. (2005). Buts d'accomplissement, stratégies d'étude, et motivation intrinsèque: Présentation d'un domaine de recherche et validation française de l'échelle d'Elliot et McGregor (2001). [Achievement goals, study strategies, and intrinsic motivation: Presentation of a research field and validation of the French version of Elliot and McGregor's scale (2001)]. *L'Année Psychologique*, *105*, 105–131. doi:10.3406/psy.2005.3821
- Darnon, C., Dompnier, B., Delmas, F., Pulfrey, C., & Butera, F. (2009). Achievement goal promotion at university: Social desirability and social utility of mastery and performance goals. *Journal of Personality and Social Psychology*, *96*, 119–134. doi:10.1037/a0012824
- Darnon, C., Dompnier, B., Gilliéron, O., & Butera, F. (2010). The interplay of mastery and performance goals in social comparison: A multiple-goal perspective. *Journal of Educational Psychology*, *102*, 212–222. doi:10.1037/a0018161
- Darnon, C., Harackiewicz, J. M., Butera, F., Mugny, G., & Quiamzade, A. (2007). Performance-approach and performance-avoidance goals: When uncertainty makes a difference. *Personality and Social Psychology Bulletin*, *33*, 813–827. doi:10.1177/0146167207301022
- DeCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York, NY: Academic Press.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, *53*, 1024–1037. doi:10.1037/0022-3514.53.6.1024
- DeShon, R. P., & Gillespie, J. Z. (2005). A motivated action theory account of goal orientation. *Journal of Applied Psychology*, *90*, 1096–1127. doi:10.1037/0021-9010.90.6.1096
- Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist*, *34*, 391–401. doi:10.1037/0003-066X.34.5.391
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*, 355–391. doi:10.1037/0033-2909.130.3.355
- Eccles, J. S. (1998). Perceived control and the development of academic motivation. *Monographs of the Society for Research in Child Development*, *63*, 221–231.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, *34*, 169–189. doi:10.1207/s15326985ep3403_3
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*, 218–232. doi:10.1037/0022-3514.72.1.218
- Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. *Educational Psychology Review*, *13*, 73–92. doi:10.1023/A:1009009018235
- Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*, 501–519. doi:10.1037/0022-3514.80.3.501
- Elliot, A. J., & Moller, A. C. (2003). Performance-approach goals: Good or bad forms of regulation? *International Journal of Educational Research*, *39*, 339–356. doi:10.1016/j.ijer.2004.06.003
- Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*, 613–628. doi:10.1037/0022-0663.100.3.613
- Elliot, A. J., & Sheldon, K. M. (1997). Avoidance achievement motivation: A personal goals analysis. *Journal of Personality and Social Psychology*, *73*, 171–185. doi:10.1037/0022-3514.73.1.171
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*, 5–12. doi:10.1037/0022-3514.54.1.5
- Garson, G. D. (2010). Linear mixed models: Hierarchical linear, random effects, multilevel, random coefficients, and repeated measures/growth models. Retrieved from <http://faculty.chass.ncsu.edu/garson/PA765/multilevel.htm>
- Goulden, N. R., & Griffin, C. J. G. (1997). Comparison of university faculty and student beliefs about the meaning of grades. *Journal of Research and Development in Education*, *31*, 27–37.
- Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, *52*, 890–898. doi:10.1037/0022-3514.52.5.890
- Grouzet, F. M. E., Vallerand, R. J., Thill, E. E., & Provencher, P. J. (2004). From environmental factors to outcomes: A test of an integrated motivational sequence. *Motivation and Emotion*, *28*, 331–346. doi:10.1007/s11031-004-2387-z
- Guttman, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). New York, NY: Free Press.
- Harackiewicz, J. M., Manderlink, G., & Sansone, C. (1992). Competence processes and achievement motivation: Implications for intrinsic motivation. In A. K. Boggiano & T. S. Pittman (Eds.), *Achievement and motivation: A social-developmental analysis* (pp. 115–137). New York, NY: Cambridge University Press.
- Harter, S. (1978). Pleasure derived from challenge and the effects of receiving grades on children's difficulty level choices. *Child Development*, *49*, 788–799. doi:10.2307/1128249

- Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education, 27*, 53–64. doi:10.1080/03075070120099368
- Hsieh, P., Sullivan, J. R., & Guerra, N. S. (2007). A closer look at college students: Self-efficacy and goal orientation. *Journal of Advanced Academics, 18*, 454–476.
- Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model comparison approach*. San Diego, CA: Harcourt Brace Jovanovich.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5*, 193–206.
- Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110*, 265–284. doi:10.1037/0033-295X.110.2.265
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284. doi:10.1037/0033-2909.119.2.254
- Koka, A., & Hein, V. (2003). Perceptions of teacher's feedback and learning environment as predictors of intrinsic motivation in physical education. *Psychology of Sport and Exercise, 4*, 333–346. doi:10.1016/S1469-0292(02)00012-2
- McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 60*, 48–58.
- McGregor, H. A., & Elliot, A. J. (2002). Achievement goals as predictors of achievement-relevant processes prior to task engagement. *Journal of Educational Psychology, 94*, 381–395. doi:10.1037/0022-0663.94.2.381
- Midgley, C., Maehr, M., Hicks, L., Roeser, R., Urdan, T., Anderman, E., & Kaplan, A. (1997). *Patterns of Adaptive Learning Survey (PALS)*. Ann Arbor: University of Michigan.
- Moller, A. C., & Elliot, A. J. (2006). The 2 × 2 achievement goal framework: An overview of empirical research. In A. V. Mittel (Ed.), *Focus on educational psychology* (pp. 307–326). Hauppauge, NY: Nova Science.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist, 22*, 155–175. doi:10.1207/s15326985ep2202_4
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328–346. doi:10.1037/0033-295X.91.3.328
- Nichols, S. L., & Berliner, D. C. (2008). Testing the joy out of learning. *Educational Leadership, 65*, 14–18.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543–578. doi:10.3102/00346543066004543
- Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology, 98*, 583–597. doi:10.1037/0022-0663.98.3.583
- Peters, R. A. (1978). Effects of anxiety, curiosity, and perceived instructor threat on student verbal behavior in the college classroom. *Journal of Educational Psychology, 70*, 388–395. doi:10.1037/0022-0663.70.3.388
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*, 544–555. doi:10.1037/0022-0663.92.3.544
- Pope, D. C. (2001). *Doing school: How we are creating a generation of stressed-out, materialistic, and miseducated students*. London, England: Yale University Press.
- Romanowski, M. H. (2004). Student obsession with grades and achievement. *Kappa Delta Pi Record, 40*, 149–151.
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*, 450–461. doi:10.1037/0022-3514.43.3.450
- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749–761. doi:10.1037/0022-3514.57.5.749
- Ryan, R. M., Connell, J. P., & Deci, E. L. (1985). A motivational analysis of self-determination and self-regulation in education. In C. Ames & R. E. Ames (Eds.), *Research on motivation in education: Vol. 2. The classroom milieu* (pp. 13–51). New York, NY: Academic Press.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78. doi:10.1037/0003-066X.55.1.68
- Ryan, R. M., & Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality, 74*, 1557–1585. doi:10.1111/j.1467-6494.2006.00420.x
- Ryan, R. M., & Grolnick, W. S. (1986). Origins and pawns in the classroom: Self-report and projective assessments of individual differences in children's perceptions. *Journal of Personality and Social Psychology, 50*, 550–558. doi:10.1037/0022-3514.50.3.550
- Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education, 7*, 224–233. doi:10.1177/1477878509104327
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144. doi:10.1007/BF00117714
- Self-Determination Theory. (2008). *Intrinsic Motivation Inventory (IMI)*. Retrieved from http://www.psych.rochester.edu/SDT/measures/IMI_description.php
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323–355. doi:10.2307/1165280
- Smith, L., Sinclair, K. E., & Chapman, E. S. (2002). Students' goals, self-efficacy, self-handicapping, and negative affective responses: An Australian senior school student study. *Contemporary Educational Psychology, 27*, 471–485. doi:10.1006/ceps.2001.1105
- Thomas, S., & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practices: "That's my grade. That's me." *Educational Psychologist, 32*, 107–123. doi:10.1207/s15326985ep3202_5
- Thorndike, E. L. (1913). *Educational psychology: Vol. 1. The original nature of man*. New York, NY: Teachers College, Columbia University.
- Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*, 389–404. doi:10.1080/0141192960220402

Received November 29, 2009

Revision received April 8, 2011

Accepted April 12, 2011 ■