The Reliability of Crowdsource Grading in a Creative Marketing Project Context

Philippe Duverger and Erin M. Steffes

Purpose of the Study: In this paper, the authors show how crowdsourcing and social media can be utilized in the creation and grading of creative student assignments without compromising academic rigor or integrity.

Method/Design and Sample: The study was conducted in three distinct steps. First, a group of students was asked to volunteer to make a creative video (n=14) describing the Millennial generation's characteristics to an audience of Baby Boomers using a social media platform. A second group of students (n=60) was asked to review the videos and evaluate each video for content and creativity. Finally, a group of four marketing educators was asked to rate the videos on both content and creativity. Inter Raters Agreement coefficients were used to test the Hypotheses.

Results: Contrary to current research (Sadler & Good, 2006), students' rating of the video projects on the basis of content was not more stringent than educators' ratings (t,20=1.08, p=.29). Students agreed with educators in both rating (IRA = .89) and ordinal ranking (z= -.27, p=.79) of the video projects on the basis of creativity.

Value to Marketing Educators: Educators can rely on crowdsourcing mechanisms to help them grade creative assignments, saving time while getting students more involved with the feedback process.

Keywords: Crowdsourcing, creativity, marketing, Web 2.0

Philippe Duverger, Ph.D. (e-mail: pduverger@towson.edu) is Assistant Professor of Marketing, Towson University. Erin M. Steffes, Ph.D. (e-mail: esteffes@towson.edu) is Associate Professor of Marketing, Towson University.

Academic learning activities that include group assignments provide students with a valuable learning experience as students interact with one another (Jacques, 2000). However, have you ever wished you could save a little time grading projects when you have few precious days to post your grades? What if you could let the students do some or all of the grading for you with the assistance of web-based technologies and social media? The concept of user-generated content (UGC) is germane to the rise of social media sites such as YouTube and Wikipedia (Kaplan & Haenlein, 2010). Educators can take advantage of social media websites to facilitate group projects, leveraging web 2.0 technologies and collaborative learning. Students are very familiar with not only the consumption of UGC but also are quite involved in its creation and diffusion. As evidence, three quarters of all Millennials have profiles on a social media website and one in five has posted a video of themselves online (Pew Research Center, 2011). Course management systems such as Blackboard now incorporate tools such as wikis and blogs that allow students to create collaborative projects and share content. It is an opportunity for a collaborative learning experience to allow student participation and collaboration in the grading process as long as the quality or integrity of the grade is not compromised.

There could be issues in the creation and evaluation of digital projects using web based technologies (i.e., slideshow presentations, videos). For example, because of the digital nature, videos cannot be printed for grading as is commonplace with final exams or term papers. Additionally, much like with a face-to-face presentation, the value of the project resides as much in the content as in the delivery mechanism and inherent creativity. Many of the grading rubrics that have been developed at universities to evaluate presentation grading include measures of content accuracy as well as presentation style; for links to sample rubrics see Appendix C. Adding to the difficulty of grading a creative digital project which is inherently subjective in nature, is the sheer quantity of submissions that could be generated if an instructor were to allow a large class to utilize these tools. Consider a Principles of Marketing undergraduate class with a 300 student roster. A sole educator could more likely grade 300 multiple choice exams easier and/or faster than 300 videos or blog posts.

The goal of this paper is twofold. First, we seek to demonstrate that there is value to educators from assigning work that contains at least an element of creativity, thus benefiting students' learning and marketability. Second, we seek to demonstrate that UGC mechanisms commonplace in social media

technology and quite familiar to students can be utilized successfully in a university setting for the creation and grading of student projects, without the loss of academic rigor or integrity by harnessing the power of crowdsourcing.

Defining Crowdsourcing

The term crowdsourcing, first introduced in Howe and Robinson's 2006 Wired Magazine article (Howe 2006), is a relatively new term that refers to the process of outsourcing the activities of a firm to an online community or crowd in the form of an "open call" (Whitla, 2009). There are many marketing crowdsourcing services available to businesses on the web; for example 99designs.com for logo design, namethis.com for brand names, and redesignme.com for product redesigns. Crowdsourcing can benefit firms because they are able to utilize a large community of potential workers with diverse skills and expertise who are willing to complete tasks in a relatively short timeframe, often at a reduced price as compared to the cost of performing the task in-house (Bayus, 2010; Brabham, 2008; Whitla, 2009). Crowdsourced tasks often include problem solving, creative thinking, and service assessment tasks (Bayus, 2010; Brabham, 2008; Whitla, 2009). Some firms have utilized the power of crowdsourcing by publishing tasks on their own websites, affording its own website traffic the opportunity to complete the task. Examples of the use and power of crowdsourcing are plentiful and include activities such as product development, creative development, and the collection of marketing research data (Whilta, 2009). An academic application of crowdsourcing introduces the concept of online collaboration into the academic journal peer review process as suggested in Lutz (2011) and Park (2010). However, contradictory opinions focus on the weaknesses and limitations of crowdsourcing; ranging from a question of expertise in the evaluating crowd to a suggestion that crowdsourcing is evil and elitist (Roman 2009, Berkus 2009).

Defining Creativity

Marketing professionals and marketing educators often struggle when asked to define creativity, but most agree that "they know it when they see it". It should not come as a surprise that finding a consistent definition of what creativity actually is can be a challenging task. There are many definitions of creativity currently in use, some focusing on the product (the output) and some focusing on the process of creation (McCorkle, Payan, Reardon, & Kling, 2007). A common theme that resonates in many of the definitions is that creativity is an activity designed to solve challenging problems in a novel manner (e.g. Amabile, 1983). A more comprehensive definition that is often used comes from Alvino (1990) who states that creative thinking is "a novel way of doing things that is characterized by four components – fluency (generating many ideas), flexibility (shifting perspectives easily), originality (conceiving of something new), and elaboration (building on others' ideas)". Finally, and what we choose as a basis for this research, Amabile (1996, p. 34) states that "for the purpose of empirical research it seems appropriate to abandon the hope of finding objective criteria for creativity and, instead, to adopt a definition that relies upon clearly subjective criteria." Amabile proposed that the most valid way to measure creativity is by using experts' judgments (1983, 1996).

THE VALUE OF CREATIVE PROJECTS

Florida and Goodnight (2005) suggest that a firm's most critical asset is "creative capital" which is defined as the "arsenal of creative thinkers whose ideas can be turned into valuable products and services". In the competitive global marketplace, many firms have adopted an organizational culture of innovation and creativity in lieu of the cost-cutting, bottom-line cultures of the 1990s (Coy, 2000). A quick review of the candidate attributes currently being sought by highly successful firms including Google, Apple, The Boston Consulting Group (BCG), and The Blackstone Group reinforces both Florida and Goodnight's and Coy's positions. Creativity shows up in hiring specifications of many firms alongside the more standard and expected candidate attributes such as leadership, integrity, ethics, independence, and initiative (Dickler, 2009). Innovation and creativity remain hot topics for businesses today well beyond the recruitment process. In addition to seeking new hires with creative skills, Fortune 500 companies are hiring "creativity consultants" to boost innovation internally (Kotz, 2011). Seventy percent of business executives responding to the 2005 BCG innovation survey stated that they planned to increase their spending on innovation (Mintz, 2006). By allowing students to develop their right brained skills via creative assignments, marketing educators can begin to align student skill sets and experiences with business needs.

Creativity in Business Schools

The importance of creativity in the marketing classroom has been addressed by a few studies (e.g., McCorkle et al., 2007, McIntyre, Hite, & Rickard, 2003; Driver, 2001; Ramocki 1994; Shipp, Lamb, & Mokowa, 1993). Recent research indicates that all business students, particularly marketing students, believe that creativity is important to their careers (McCorkle et al., 2007). Creative activities such as acting, dancing, improvisation, and musical performance have begun to infiltrate the traditionally leadership and analysis-focused MBA programs in top universities including the Darden School of Business, MIT Sloan, and Butler University. These atypical MBA learning experiences are designed to help students develop the typically right brained skills by utilizing their creative imagination (Feinberg, 2012). In deference to the level of creativity and innovation required for success in today's competitive marketplace, business schools are

reviewing curriculum across the United States to determine where creativity courses could be added and/or where creativity can be infused into current classes and curriculum (Redesigning the MBA, 2012). To meet marketplace and student needs, many educators are seeking ways to add projects and assignments to their classes that can foster creativity and get students thinking "outside of the box". However, lingering concerns about creative projects remain. A recent focus group on creativity in the classroom at a large metropolitan university highlighted two primary concerns over the focus on creativity in a business school curriculum. The first concern raised by the group was whether creative assignments would come at the expense of content delivery from the educator. The second concern was whether faculty possessed the time and the knowledge to effectively evaluate and grade the creativity of the projects (CBE LEC focus group, November 30, 2011). While faculty felt confident that they could evaluate the subject matter content for any project, they expressed concerns about their ability to evaluate the creative elements. Nussbaum (2005) suggests that while business schools have made some progress in incorporating creativity into the business classroom, he finds that for the most part, business schools do not really understand creativity nor do they teach it very well. In fact, popular press goes as far as to suggest that business schools can extinguish students' creative spark (Stillman, 2011).

Using Social Media to Facilitate Creative Marketing Projects

Marketing educators may need to rethink pedagogy for the digitally native Millennial generation. The learners of today will be called upon in the workforce to not only be consumers of information but also to be co-producers of information in their socially connected digital world (McLoughlin & Lee, 2007). To evolve students beyond passive content consumption to active content creation will require that business students expand their skill set to include creative skills. One way to increase the level of creativity in marketing classrooms is to introduce assignments and projects that differ from what one might deem as the more "traditional assignments" such as papers and tests. Many marketing classes still focus student effort and evaluation on traditional assignments even though the typical college-aged student, the Millennial, gains knowledge and consumes media in a very different manner than previous generations. Many Millennials consider their use of technology, including the Internet, to be the defining characteristic of their generation; 24% of Millennials state that their use of technology is what makes their generation unique (Krigman, 2010). A broad definition of social media proposed by Kaplan & Haenlein (2010) suggests that "social media includes a group of Internet-based application that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content." While only 7% of Millennials used social networking sites in 2005, 75% of Millennials used social networking sites in 2010 (Krigman, 2010). Millennials' social networking usage penetration far exceeds usage penetration in older generational cohorts. Millennials have rapidly adopted the use of many social media sites such as Facebook, Wikipedia, Second Life, YouTube, and Twitter.

Social media provides marketing educators with a platform to utilize in the assignment of creative projects. A content community's main objective, aside from profit making, is facilitating the sharing of content between users; YouTube, for example, allows users to post and share user-created digital videos with ease (Kaplan & Haenlein, 2010). Marketing educators can assign the simple task of creating an educational video teaching a marketing concept such as perceptual mapping without encumbering students with content and format restrictions and detailed instructions typically found in written assignments such as homework, case analyses, and term papers. Students are then free to let their creativity take center stage as they plan and create digital content. Part of the beauty of this type of assignment is that it works well for either individual projects and for group projects.

Crowdsourcing: Using Social Media to Grade Creative Marketing Projects

An academic social media application of crowdsourcing could operate in a fashion similar to the commercial Web 2.0 because the very nature of crowdsourcing captures the essence of user generated content. After digital student projects are created and submitted utilizing social media such as YouTube, the educator can use crowdsourcing to grade the project. If the course was supported by a learning management system such as Blackboard, links to the YouTube videos could be posted directly into the course site. The videos could be set as private on You Tube, thus protecting the students; yet, links would be shared within Blackboard for grading. Students could then be asked to review their peers' work by either posting comments directly onto Blackboard to provide feedback on the videos, or by completing a simple evaluation survey hosted by Blackboard or any of the free online survey websites such as surveymonkey.com. To facilitate the crowdsourcing process, technologically savvy educators might choose to embed YouTube links directly into an online survey that would then be emailed to the student roster for review and completion. It would be at the discretion of the educator as to whether the students would be asked to evaluate content, creativity or both. In fact, crowdsourcing has already seen some application in college classrooms via blogs with a rubric-based grading method that holds students more accountable to the responsibility of becoming contributors in the evaluation of peer work (Hendry 2009).

HYPOTHESES DEVELOPMENT

Involving students in the assessment of student work can take two forms: peer assessment or self-assessment. The focus of our study is on the use of peer assessment; when students assess the work of other students; rather than their own.

Peer assessment is grounded in the philosophy of active learning (Piaget, 1971), and may also be a manifestation of constructionism (Vygotsky, 1962) since it involves the joint construction of knowledge through discourse (Falchikov & Goldfinch, 2000). Educators might avoid peer assessment for fear of lack of reliability or validity. The present study addresses the concern of validity (does the students' assessment match that of the teacher's) and also addresses reliability (do student assessments match that of other students) as a necessary but not sufficient condition.

In their meta-analysis, Falchikov & Goldfinch (2000) reviewed 48 quantitative peer assessment studies and concluded that peer assessments resemble more closely teacher assessments when global judgments are made using fewer rather than larger dimensions. Peer assessment matched educator assessment better in purely academic studies rather than in professional studies. The authors also suggest based on their analysis, that more peer raters are not better than few in matching educator assessments. The authors calculate an average mean correlation of .69 over all studies.

In their research on how student learning might be affected by peer-grading, Sadler & Good (2006) found that, despite a very high correlation (r=.91 to .94) between the educator's and students' grades, some bias existed. When grading their peers, students awarded significantly lower grades to the best performers than the educator did. Furthermore, research found that students had more difficulty grading items testing higher order understanding such as creative work (Zoller, Tsaparlis, Fastow, & Luberesky, 1997). Thus we propose:

> **H_1:** Students' mean rating of the assignment's creativity will be significantly lower than educators' mean rating.

While creativity is the main thrust of the study, we felt content needed to be measured as well. Not only would content constitute a form of technical goodness (Amabile, 1984, 1996 p. 46) that could be correlated to creativity; content would also be a good base level measure across experts. Thus we propose:

> **H_2:** Students' mean rating of the assignment's content appropriateness will be significantly lower than educators' mean rating.

Sadler & Good (2006) found that despite the lower grade awarded to peers students' grades on average were highly correlated with the educator's grades. Hence, these findings suggest that ranking might not be different between educators and students, although Sadler & Good (2006) do not report rankings. Thus we propose:

> **H_3**: Students' ordinal ranking of the assignments on the basis of creativity will not be different than educators' ranking.
> **H_4:** Students' ordinal ranking of the assignments on the basis of content appropriateness will not be significantly different than educators' ranking.

METHOD

The study was conducted in three distinct steps. First, a group of students was asked to volunteer to make a creative video explaining the Millennial generation's characteristics to an audience of Baby Boomers (See Appendix A). We used videos as assignments but conceivably one could use other creative media such as animation, collages, podcasts and so on. A second group of students was asked to review the videos and rate the videos individually for both content and creativity. All students were volunteers and received extra credit for their involvement. Those who did not choose to participate in the video creation or assessment project were given an alternative extra credit project. Finally, a group of four marketing educators all tenured or on tenure–track, was asked to also rate the videos on both content and creativity.

Video Assignments

Out of a class of thirty-five, eleven male and nine female students participated in the creation of fourteen videos with an average length 2.24 minutes (SD=1.06) (participants had the choice of working individually or pairing up for the project). The assignment asked that the participants make a video with a maximum length of 5 minutes and encouraged them to be creative in communicating their message. Participants were not given specific instructions on the required content or the format for the video, since the project resembled a typical short answer exam question and students had been exposed to the material during the class.

Once the videos were created, the students were asked to upload the videos to a private YouTube account and send the link to one of the researchers via Blackboard. The videos ranged widely in the use of media, music, settings, and overall context allowing the researchers to anticipate sufficient variability in the creativity ratings. Similarly, the videos' content showed variance in the application of segmentation theory and use of examples to convey the required information to the Baby Boomer target audience. The videos as a whole were eclectic; ranging from a self-filmed student reading the text to a true mash-up of slides, pictures, voice-over and text both explaining and showing the unique characteristics of the Millennial generation.

Video Assessments using the Consensual Assessment Technique

First proposed by Teresa Amabile the Consensual Assessment Technique (CAT) is a validated tool for

assessing creativity (Amabile, 1982, 1996; Baer & McKool, 2009). Amabile (1982) found that other attributes could be measured using the CAT process including less creative and more technical attributes; although Amabile found weaker agreement among experts on the non-creative attributes being measured. In the basic CAT procedure the subjects (in our case the students) are given basic instructions to create a product (videos). Next, a group of experts (e.g., the marketing educators, the student crowd), who are knowledgeable of their domain (Millennial segmentation) worked independently of one another, to judge the creativity of those products. The judges are not asked to explain or defend their ratings in any way, and it is important that no such instructions be given. Judges are simply asked to use their expertise in assessing the creative products in relation to each other. Products are not compared to an ideal creation, but are compared within the relative pool of product being judged. Judgments on all dimensions, creativity and content, are made at the same time. With CAT products must be rated in a random order (Baer & McKool, 2009; Starko, 2010).

The present research asked the judges (the marketing educators- the experts, and the students- the crowd) to evaluate the videos on the basis of the Consensual Assessment Technique (CAT).

Prompted by the following guideline (See Appendix): "Using your own subjective definition of creativity", the judges had to answer the following statements (using a 7-point Likert scale – anchored by Strongly Disagree (1)-Strongly Agree (7)) adapted from Amabile (1982, 1996, p.46):

1. This video is original
2. This video is creative
3. This video used unique ways to make its point

Three questions were asked of judges to assess the content of each video (using a similar 7-point Likert scale):

1. This video explains what segmentation is very well
2. This video teaches me the key characteristics that differentiate Millennials from Baby Boomers
3. This video gives good examples of the segments' differences

Video Assessment by "the Experts"

Four marketing educators, with marketing professional and teaching experience of at least 10 years, were asked to rate the videos. Previous research using the CAT used anywhere from 2 to 40 judges (Baer & McKool, 2009). Each educator rated all 14 videos. Instructions for the experts were simply to watch the videos and answer the questions (See Appendix B). Raters were informed of the tasks that were to be accomplished by the students, and they also knew the course these students were enrolled in, hence they could infer the level of knowledge these students have. The process of rating for all marketing experts took less than 48 hours and was done during the same week that the students performed their ratings.

Video Assessment by "the Crowd"

Students majoring in marketing from a metropolitan university located in the central east cost of the US were invited to participate in the rating of the videos. In this study the marketing students represents the "crowd". We take the definition of a crowd to be any one who responds to an open call (Whitla, 2009) thus any marketing major students who volunteered to rate the videos instead of completing the alternate assignment. The students represent the "marketing student crowd" and not the university at large. Given that the marketing students should have the relevant subject matter knowledge, the study seeks to establish that they possess, as a group, the expertise to rate content as well.

Students were given 5 points extra credit as a motivational mechanism to encourage participation in the experiment. The provision of an incentive is not uncommon in crowdsourcing. Sixty students from two classes of 35 students each participated in the rating. The sample reflected the characteristics of the student body and was composed of juniors (29.7%) and seniors (70.3%) with 62.3% female students, and a self-reported mean GPA of 3.75. The videos were embedded in an online survey so that each video would automatically play, with the questions presented simultaneously with the video. The video display and the questions mirrored the look and feel of You Tube format. To avoid fatigue each student was presented with no more than 6 videos out of the total of 14, totaling less than 8 minutes of viewing. The online mechanism allowed for a complete randomization of the process and each video was rated by a minimum of 20 students and a maximum of 23. The entire process of online video rating was open for student participation for one week. Students took on average 10.3 minutes to review and rate their videos. The students used the same criteria as the educators to rate each of the videos presented to them.

Intraclass Correlation Coefficients

The Intraclass Correlation Coefficients (ICC) differs from the other bivariate correlation coefficients such as the Pearson correlation coefficient, in that the general goal with ICC is to determine the proportion of variation in a measure that is due to being a member of a particular group called a class. Partition of the variance follows an ANOVA method which boils down to an understanding the sources of systematic variance in the ratings.

There are two broad categories of ICC: the first seeks consistency in the ratings, also known as the interrater reliability (IRR) which is a measure comparable to the well-known Cronbach's Alpha; the other seeks to demonstrate the agreement between raters and across classes, known as the interrater agreement (IRA) measure (McGraw & Wong, 1996; Shrout & Fleiss, 1979).

The main difference between IRR and IRA resides in the formula largely explained in McGraw & Wong (1996). In the case of the IRR, the variance between raters is not factored into the calculation and the goal is to find whether the object evaluated is consistent in its relative ranking based on the scores given. In the case of the IRA, variance between raters is factored in since the objective is to find out whether or not the raters agree not only on the ranking but also on the score given to the object evaluated.

We will first focus on the IRR, i.e., do the raters consistently rank the videos the same way, based on their composite scores. For example, if a video differs in absolute score, but still ranks first between two raters, the IRR will be high (i.e., close to 1). A high level of consistency will allow for inference at the group level (educators versus students) and will permit the aggregation of all raters' scores within each group. Ratings can be averaged across raters (Rust & Cooil, 1994) once the reliability (IRR) is equal to or greater than .70 (Boulding, Staelin, Zeithaml & Kalra, 1993).

We will use IRA to determine if the students, as a group, gave the videos the same scores as the educators as a group. Agreement is said to be demonstrated if the IRA is greater than .70 (McGraw & Wong, 1996; Shrout & Fleiss, 1979), although some recent research calls for a higher threshold of .90 (Harvey & Hollander, 2004). Inferences on both the IRR and the IRA can be conducted using an F-test, and one can infer whether the IRR or the IRA have meet a threshold equal to an expected effect size (Cohen, 1988; McGraw & Wong, 1996).

RESULTS

Three of the videos did not play reliably during the experiment and could not be rated by educators; hence they were removed from the analysis. This does not constitute a limitation since the intent of the study is to find the power of the crowd in agreeing with the experts, rather than determining which particular video ranked first.

Reliability of the Constructs Measures within the Groups of Raters

For each video a minimum of 20 student answers per question was collected and reliability of each construct (content and creativity) was assessed using Cronbach's Alpha (Nunnally, 1978). Similarly, the four educators rated the videos, and as can be seen on Table 1 both educators' and students' ratings present a high internal reliability for each construct and each video. The Cronbach's Alphas range from .66 to a high 1.0.

Table 1: Content and Creativity Constructs' Cronbach's Alpha per Video

Video	Construct	Alpha Students[a]	Alpha Educators[b]
1	Content	0.83	0.88
	Creativity	0.89	0.89
2	Content	0.82	0.89
	Creativity	0.85	0.88
3	Content	0.79	0.73
	Creativity	0.80	1.00
4	Content	0.75	1.00
	Creativity	0.84	0.84
5	Content	0.89	1.00
	Creativity	0.94	0.92
6	Content	0.66	0.75
	Creativity	0.96	0.96
7	Content	0.93	0.76
	Creativity	0.91	0.99
8	Content	0.88	1.00
	Creativity	0.97	0.73
11	Content	0.84	0.89
	Creativity	0.93	0.96
12	Content	0.79	0.92
	Creativity	0.69	0.92
14	Content	0.89	0.88
	Creativity	0.93	0.92

Note: videos 9, 10, and 13 omitted due to a technical

playback problem and lack of evaluation.
a. IRR $\rho_{content\text{-}students}$=.993; $\rho_{creativity\text{-}students}$=.990.
b. IRR $\rho_{content\text{-}educators=}$.672; $\rho_{creativity\text{-}educators=}$.743.

These results allow us to be confident that each question measures the same construct and therefore aggregation by averaging the scores is permissible. Hence, for each video and each construct, an average score was calculated representing the composite score per respondent. Thus, each video now has two composite scores (content and creativity) for each of the students who rated the video and each of the educators.

Interraters' Consistency

For each group of raters, both student and educator, an IRR was calculated (Shrout & Fleiss, 1979). Because a sub-set of students was drawn randomly form a larger pool (70 students) to evaluate a set of 6 videos a two-way random IRR is the appropriate measure to use for consistency (McGraw & Wong, 1996) within the group. This measure is generalizable to all possible judges (Shrout & Fleiss, 1979). For the student group the average IRR measure showed a strong consistency among the students ($\rho_{content}$=.993 and $\rho_{creativity}$=.990). Hence, all students' scores on content and creativity can be aggregated together make one content and one creativity score per video.

Similarly we obtain an average IRR measure for the educators' group of .672 for content and .743 for creativity. These IRR results are substantial yet not outstanding as one would prefer IRR > .90 (Harvey & Hollander, 2004). However an F-test associated with the IRR coefficient reveals that in their population educators would only rate consistently in approximately 5 cases out of 10 on content, and 4 out of 10 on creativity[1]. Hence, even if the population's IRRs are both different from zero the evidence seems to suggest that educators among themselves rate significantly differently as compared to the students.

The purpose of the present study is less to infer population IRRs than it is to look for a potential agreement between the educators and students rating in the sample. Hence, we turn our analysis to the hypotheses testing.

Hypotheses Testing

Based on the IRRs obtained in the sample we aggregated all students' scores for each video on the construct content, yielding one average content score per video. We did the same with the students' creativity scores. We then aggregated the educators' scores to calculate the aggregated scores shown in Table 2.

[1] F-test using McGraw & Wong (1996) method with an hypothesized value for IRR matching the critical value for F of 2.3479 yielded .23 for the content IRR and .39 for the creativity IRR.

Table 2. Aggregated Scores per Video

Video	Content Educators a	Content Students b	Absolute Difference a-b	Creativity Educators c	Creativity Students d	Absolute Difference c-d
1	4.500	3.778	0.722	2.889	2.875	0.014
2	5.667	4.651	1.016	3.111	4.048	(0.937)
3	5.500	4.826	0.674	2.333	3.087	(0.754)
4	5.667	5.591	0.076	4.833	4.591	0.242
5	5.556	5.136	0.420	4.444	4.545	(0.101)
6	6.333	5.870	0.463	5.111	5.058	0.053
7	4.222	4.848	(0.626)	4.778	4.167	0.611
8	6.333	5.576	0.757	5.500	5.258	0.242
11	5.222	4.847	0.375	3.556	3.375	0.181
12	4.222	4.097	0.125	2.778	2.806	(0.028)
14	4.000	4.136	(0.136)	4.222	2.727	1.495
Mean	5.202	4.850	0.352	3.959	3.867	0.092
S.E.	0.25	0.20	0.32	0.32	0.28	0.43

To test our first two hypotheses we began by conducting an independent-sample t-test of means applied on the aggregated video scores listed in Table 2. Students' mean rating of creativity is lower than educators' mean rating (educator $mean_{creativity}$=3.96, SE=.32; student $mean_{creativity}$=3.87, SE=.28), however

not significantly ($t_{(20)}$=.22; p=.83) thus rejecting H1 in the population. Students' mean rating of content is not significantly lower than educators' mean rating (educator $mean_{content}$=5.20, SE=.25; student $mean_{content}$=4.85, SE=.20; $t_{(20)}$=1.08; p=.29), thus H2 is not supported.

Using the results shown in Table 2 we calculated a one-way IRA of .846 for content and .894 for creativity. Confidence intervals respectively equal .456 and .958 for content, and .628 and .971 for creativity. Furthermore an F-test with an alpha threshold of .05 yields an IRA of .67 for creativity in the population and .70 for content; a medium effect size according to Cohen (1988). The average IRA is greater than .70 but still shy of the recommended threshold of .90 (Harvey & Hollander, 2004); furthermore, the lower bound of the confidence intervals would suggests that students and educators vary in their rating of content quite more in the population than they would vary when evaluating creativity. Thus, even if the agreement on content rating in the sample is high, it could be as low as .456 in the population. Therefore, we can be confident that in our sample the students agree with the educators on the rating of the videos, yet further studies should investigate if the weakness implied by the lower bound of the confidence internal is a reflection of the population's distribution, a factor of our sampling, or a factor of assessment skills. For instance, some research shows that interrater reliability can be increased by using rubrics (Sadler & Good, 2006). These results are consistent with other research (Amabile, 1996) where technical assessment agreements were often weaker than creativity assessments.

Video Ranking

Given the confidence intervals of the IRA calculated above, the aggregate scores sorted in decreasing order yield an ordinal ranking from the "best video" based on content and creativity to the worst video based on the scores given by educators and students (See Table 3).

Table 3. Rankings per Video

Video	Content Educators	Content Students	Creativity Educators	Creativity Students
1	8	11	9	9
2	3	8	8	6
3	6	7	11	8
4	4	2	3	3
5	5	4	5	4
6	1	1	2	2
7	10	5	4	5
8	2	3	1	1
9	7	6	7	7
10	9	10	10	10
11	11	9	6	11

Non-parametric Wilcoxon signed rank non-directional tests were conducted on the data in Table 3. The ranking of videos by educators and students based on the raw scores are not significantly different for creativity (z=-.271, p=.786), or content (z=-.103, p=.918) providing support for both H3 and H4.

SUMMARY OF FINDINGS AND DISCUSSION

Both popular press and academic literature suggest that creativity should have a more prominent place in U.S. business schools. Hiring firms seek graduates with both a solid understanding of business theory and the ability to think creatively as employees address the firm's challenges. However, business educators are often hesitant to integrate creative assignments into their courses due to the inherent subjectivity and difficulty in evaluating and grading creative assignments. Marketing educators have many opportunities to integrate creativity into their courses because marketing is both quantitative and qualitative in nature. Social media integrated with course delivery systems provide educators with a forum for the submission and grading of creative assignments through platforms including YouTube and Twitter. Creative marketing assignments such as mock television commercials, personal selling role play, and new product pitches could easily be incorporated into marketing courses allowing students to show their mastery of the subject matter while letting their creativity shine through.

The present study utilized a fairly simple and straightforward assignment that asked students to teach Baby Boomers about the Millennial

segmentation characteristics by creating a YouTube video. This type of YouTube "teaching" assignment could be readily adapted for various audiences and subject matter. A roster of students could be asked to choose one topic from the course and create a video teaching or illustrating the topic of their choice to their peers. The assignment works well as either an individual or a group project. Once the YouTube video links are submitted to the educator, the videos could be made available to the class through the course management software in order to protect privacy. This library of videos could then serve as a ready reference source for the class on various marketing topics.

Many educators have come to understand that most students will not exert significant effort on any assignment that does not receive a grade. Therefore, grading and feedback play a critical role in integration of creative assignments into marketing courses. We show how crowdsourcing can be used to help alleviate the burden of grading creative assignments. In the present study we hypothesized that the "crowd" i.e. the marketing students, would agree with the educators in ranking of video projects on the basis of content and creativity, but that they would be more harsh in the actual rating, giving the videos a "grade" significantly lower than the educators.

Students seem more stringent in rating the video projects on the basis of content appropriateness and give lower ratings than educators, although not in a statistically significant manner. Yet these lower rating do not yield significantly different ordinal rankings. Students do agree with educators in ranking video projects on the basis of content and creativity. In fact, students and educators rank the top 3 videos (out of 11) for creativity in perfect agreement.

The somewhat lower rating given by students is in line with previous research and might be remediated by the use of rubrics (Boud, 1989; Sadler & Good, 2006). An educator that would want to use the ranking as a basis for grading could trust the "crowd" ranking to not be different than his/her own and assess grades on the basis a predetermined curve based on ranking. In this manner, the educators would not have to correct for the students' rating bias.

It is therefore our recommendation that crowdsourcing be used in the evaluation of creativity thereby eliminating the more challenging grading element for creative projects, while engaging the students in an activity that mirror their own social media daily activities.

The results of this study have some limitations. First, the participants were marketing students, and therefore future study should investigate the appropriateness of crowdsourcing with other disciplines within business majors, and beyond. Second, we did not measure likeability as a construct. Research has shown that peer-assessment can confuse likeability for creativity although the research participants were younger students (Hickey, 2001). Finally, for the crowdsourcing mechanism to work well in the context of grading, all videos need to play back well. Therefore the technology supporting a crowdsourcing process should be robust and all involved should be trained on its use and able to trouble shoot some of the inevitable problems; or be ready to revert to more traditional methods.

Several limitations can undermine the generalizability of the present study. First, other discipline than marketing could yield different results. Second, the standing of the students' experts (juniors and seniors) could also have had an effect on the level of agreement reached with the faculty experts, and freshmen or sophomores might not have the same expertise level. Third, we chose videos as the creative assignment medium, but others might prefer podcast or collages, and the medium used might imply easier or more difficult assessments, or in the case of a live presentation, might induce bias due to the possible friendship between students experts and students creators. Finally, students self-selected in the creation of the videos, therefore the sample of video is possibly skewed toward more creative videos. This could be a problem in the aggregate if our purpose was to evaluate class averages or college averages, however in our case, the aim of the study being agreement between experts; the self-selection bias canceled itself out. Similar bias is present in the expert crowd; here the bias could be an alternative explanation of agreement, although the all crowdsourcing based project suffer from the same limitation.

REFERENCES

Resigning the MBA: A Curriculum Development Symposium (2012). retrieved 1/27/ 2012 from http://www.aacsb.edu/MBASymposium/index.html.

Alvino, J. (1990). A glossary of thinking-skills terms. *Learning,* 16, 50.

Amabile, T. M. (1982). Children's artistic creativity: detrimental effects of competition in a field setting. *Personality and Social Psychology Bulletin*, 8: 573-578.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 43: 357-376.

Amabile, T. M**.** (1996). *Creativity in Context*. Boulder, CO: Westview Press.

Baer, J., & McKool, S. S. (2009). Assessing Creativity Using the Consensual Assessment Technique. In C. Schreiner (Ed.), *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education* (pp. 65-77). doi:10.4018/978-1-60566-667-9.ch004

Bayus, B.L. (2010). Crowdsourcing and individual creativity over time: the detrimental effects of past

success, *working paper* accessed at http://ssnn.com/abstract=1667101

Berkus, J. (2009). Never say crowdsourcing. Toolbox.com, September 2009. Retrieved 5/15/12 from http://it.toolbox.com/blogs/database-soup/never-say-crowdsourcing-34331

Boud, D. (1989). The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education*, 14, 20-30.

Boulding, W., Staelin, R., Zeithaml, V. & Kalra, A. (1993). A dynamic process model of service quality: From expectations to behavioral intentions, *Journal of Marketing Research*, 30:7-27.

Brabham, D.C. (2008). Crowdsourcing as a model for problem solving: an introduction and cases, *Convergence: The international Journal of Research into New Media Technologies*, 14(1): 75-90.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Coy, P. (2000, August 28). The 21st century corporation: The creative economy. *Business Week,* 77-82.

Dicker, J. (2009). Top 15 MBA Employers. CNN Money, 5/27/2009. Retrieved 1/16/2012 from http://money.cnn.com/galleries/2009/fortune/0905/gallery.top_mba_employers.fortune/9.html

Driver, M. (2001). Fostering creativity in business Education: developing creative classroom environments to provide students with critical workplace competencies. *Journal of Education for Business*, 77 (1): 28-33.

Falchikov, N. & Foldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, 70(3): 287-322.

Feinberg, S. (2012) Creativity in the class. *BizEd*, March/April 2012, 43-46.

Florida, R. & Goodnight, J. (2005). Managing for creativity. *Harvard Business Review*, 83:124- 132.

Jacques, D. (2000). Learning in Groups, (3rd Edition). Kogan Page Lt. London.Joining Google (2012).. Retrieved 1/17/2012 from http://www.google.com/intl/en/jobs/joininggoogle/index.html

Harvey, R.J. & Hollander, E. (2004). Benchmarking rWG Interrater Agreement Indices: Let's Drop the .70 Rule-Of-Thumb in *Annual Conference of the Society for Industrial and Organizational Psychology*.

Hendry, E. (2009). Duke professor uses 'crowdsourcing' to grade. Chronicle of Higher Education. July, 2009. Retrieved 5/15/12 from http://chronicle.com/blogs/wiredcampus/duke-professor-uses-crowdsourcing-to-grade/7538

Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education*, 49(3): 234-244.

Howe, J. (2006). The Rise of Crowdsourcing. *WIRED Magazine* (14.06), accessed at http://www.wired.com/wired/archive/14.06/crowds.html

Kaplan, A.M. & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media, *Business Horizons*, 53: 59-68.

Kotz, D. (2011). 6 Steps to achieving creativity in business, personal life. *US News.* Retrieved 1/17/2012 from http://health.usnews.com/health-news/family-health/brain-and-behavior/articles/2011/01/03/6-/steps-to-achieving-creativity-in-business-personal-life

Krigman, E. (2010). Millennials defined by technology use. *National Journal.* Retrieved 1/17/2012 from http://www.nationaljournal.com/njonline/no_20100225_3691.php

Lutz, R. (2011). Marketing Scholarship 2.0. *Journal of Marketing,* 75 (4): 225-234.

McCorkle, D., Payan, J., Reardon, J., & Kling, N. (2007). Perceptions and reality: Creativity in the marketing classroom. *Journal of Marketing Education*, 29: 254-261.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1): 30-46.

McIntyre, F. S., Hite, R. E., & Rickard, M. K. (2003). Individual characteristics and creativity in the marketing classroom: Exploratory insights. *Journal of Marketing Education*, 25: 143-149.

McLoughlin, C., & Lee, M.J.W. (2007). Social software and participatory learning: pedagogical choices with technology affordances in the Web 2.0 era, in ICT: providing choices for learners and learning, *Proceedings ascilite Singapore 2007*, available at: www.ascilite.org.au/conferences/singapore07/procs/mcloughlin.pdf

Mintz, P. (2006, March 16). Creativity comes to Business schools. *Business Week*, Retrieved January 16, 2012 from http://www.businessweek.com/bschools/content/mar2006/bs20060326_8436_bs001.htm

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Nussbaum, B. (2005, January 3). Getting schooled in innovation. *Business Week*. Retrieved 1/18/2012 from http://www.businessweek.com/bwdaily/dnflash/jan2005/nf2005013_8303.htm

Pew Research Center (2011). The Millennials: Confident. Connected. Open to Change. Retrieved on5/15/2012 from http://pewresearch.org/pubs/1501/Millennials-new-survey-generational-personality-upbeat-open-new-ideas-technology-bound

Park, C.W. (2010). A Two –Step Interactive Review Process. *Journal of Consumer Psychology*. 20 (2):105-106.

Piaget, J. (1971). *Science of Education and the Psychology of the Child*. Oxford: Pergamon Press.

Ramocki, S. (1994). It is time to teach creativity throughout the marketing curriculum. *Journal of Marketing Education*, 16:15-25.

Roman, D. (2009). Crowdsourcing and the question of expertise. *Communications of the ACM*, 52 (12). Retrieved on 5/15/2012 from http://dl.acm.org/citation.cfm?id=1610258&bnc=1

Rust, R. & Cooil, B. (1994). Reliability measures for qualitative data: Theory and implications. *Journal of Marketing Research*, 31:1-14.

Sadler, P.M. & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment*, 11(1): 1-31.

Shipp,S., Lamb, C. W. & Mokowa, M. P. (1993). Developing and enhancing marketing students' skills: Written and oral communication, intuition, creativity and computer usage. *Marketing Education Review*, 3, 2-8.

Shrout, P.E. & Fleiss, J.L. (1979) Intraclass correlations: users in assessing rater reliability. *Psychological Bulletin*, 86(2): 420-428.

Starko, A.J. (2010). *Creativity in the Classroom: Schools of Curious Delight*, 4th Ed. NY: Routledge.

Stillman, J. (2011). Will business schools extinguish your spark? *Inc.* Retrieved 1/16/2012 from http://www.inc.com/jessica-stillman/will-business-school-extinguish-your-spark.html

Vygotsky, L.S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Whitla, P. (2009). Crowdsourcing and its application in marketing activities, *Contemporary Management Research*, 5(1): 15-28.

Zoller, U.; Tsaparlis, G.; Fastow, M. & Luberesky, A. (1997). Student self-assessment of higher-order cognitive skills in college science teaching. *Journal of College Science Teaching*, 27, 99-101.

Appendix A: Instructions for the Video Production

For credit (5 points)
Produce a short video (less than 5 minutes) on the following subject: Describe the Market Segmentation of the Generation Y (note: Gen Yers are born after 1982)

- The video has to target Baby Boomers (born 1946 and 1964)
- The video has to be as lively and as creative as possible.
- Mash-ups of existing content are acceptable, so is the use of music, and text, with or without citations.
- Videos using strictly stock picture, or re-cycling (plagiarizing) entire video content will be rejected.

Once finished upload your video to You Tube (using your account, but set the video as private) and post the link to Blackboard. Do not forget to include your name and class in order to get the credit.

Appendix B: Instructions for the Video Assessment

You will see X videos presented to you in no particular order.
Your job is to analyze each video on two main criteria:
1/ How well the video explains the Gen Y segmentation (age 18-28) characteristics to the Boomer generation (50 years old +). By characteristics we include demographics (age, gender, ethnicity, income) and psychographics (habits, personality traits, behaviors, beliefs)
2/ Using your own subjective definition of creativity, how creative and engaging the video is to you?

For each video, please:
1/ Watch the video (make sure your sound system is working)
2/ Answer the questions
3/ Go to next video
Thank you

Appendix C: Links to Sample Rubrics

http://www.ncsu.edu/midlink/rub.pres.html,
http://ed.fnal.gov/lincon/w01/projects/library/rubrics/presrubric.htm
http://www.readwritethink.org/files/resources/lesson_images/lesson416/OralRubric.pdf