

INSTRUCTIONAL HUMOUR AND COGNITIVE AFFECTIVE LEARNING WITH MULTIMEDIA (IHCALM)



SCAN ME

Diedon DORAMBARI 

Independent Scholar, d_diedon@hotmail.com

Article history:

Submission 03 December 2021

Revision 29 January 2022

Accepted 03 March 2022

Available online 30 April 2022

Keywords:

CATLM,

Instructional Humor,

IHCALM.

DOI:

<https://doi.org/10.32936/pssj.v6i1.282>

A b s t r a c t

While background music and interesting yet irrelevant to the topic adjuncts were found to harm learning (and were classified as seductive detail) in the Cognitive-Affective Theory of Learning with Media model (CATLM), emotionally appealing shapes and color were found to foster learning (and were classified as multimedia with emotional design). However, although humour is used in education during class and has both psychological and physiological benefits, there is no published research about instructional humour (IH) in CATLM to date. The purpose of the current research was to clarify whether IH in CATLM fosters learning, or if it is yet another type of seductive detail. Total of 96 young undergraduate student participants were randomly assigned to watch a stimuli depicting 3D animations of brain cells either with IH (named as IHCALM) or without it (named as NH). All student data regarding mirth duration were measured with cameras, while how funny they found the stimuli, as well as their cognitive load, emotions, motivation, knowledge, and metacognition were all measured with OpenSesame. To test if the IHCALM harms learning, similarity between conditions was analyzed with both Bayesian Factor analysis and null hypothesis testing, which jointly reveal 3 outcomes. Outcome results show that IHCALM does not harms learning, due to being similar with the non-humorous condition. Implications of these findings for education are considered.

1. Introduction

Humor involves both the cognitive and reward components (Franklin & Adams, 2011; Vrticka et al., 2013), as well as intelligence and creativity (Greengross et al., 2012). As such, the use of humor in education is believed to have a significant history (Wilkins & Eisenbraun, 2009). When humor is used in education, the learning process becomes a joint educational growth-inciting venture for both the teacher and the learner (Hackathorn et al., 2012; Morrison & Quest, 2012). However, although the cognitive affective theory of learning with media (CATLM, Moreno & Mayer, 2007) is evolving to include various new emotion inciting elements (Um et al., 2011), there is no published research to date that includes instructional humor (IH) as an IV in CATLM.

Largely due to concerns related to working memory limitations, introducing a new type of instruction (e.g., IH) as an IV to test the CATLM DVs remains a challenge with multimedia designers to date (Mayer & Estrella, 2014). For instance, when the IV is

instruction and emotionally appealing adjuncts, spectacular videos, or soothing background music, then the incited emotions impeded learning with the “seductive detail” effect. The seductive detail effect harms learning because it unnecessarily extrinsically cognitively loads the learners’ limited working memory units with non-intrinsic instruction (Harp & Mayer, 1998; Park et al., 2015). As such, it is not clear if IH as an IV would also needlessly overload the learners’ limited cognitive resources (i.e., result in yet another type of “seductive detail”), or would it foster learning genuinely instead.

If IH does not harms learning (i.e., is not another type of seductive detail), then the results of tests after watching a multimedia presentation with IH should be similar to the one without it. However, since absent of evidence does not means evidence of absence, then in addition to null hypothesis testing, Bayesian analysis was also used to measure the degree the two conditions

would be similar. In this study, the results between the two conditions was found to be similar on most of the CATLM DV's.

2. Objective

Thereby, the study aimed to test if humor aids or harms learning in CATLM. To do so, humor will be used as an IV to incite most of the DVs mentioned in the CATLM model, such as cognitive load, academic emotion, motivation, learning, and (for the first time in cognitive multimedia learning) metacognition. The importance of these DVs and their relationship are depicted in Figure 1.

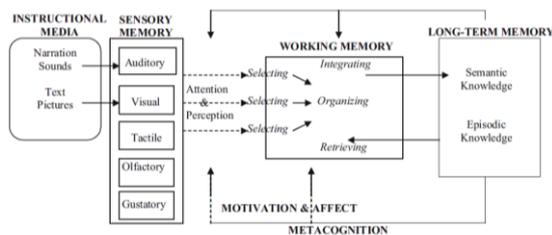


Figure 1. The cognitive-affective theory of learning with media
Source: Moreno and Mayer (2007).

3. Method

The IH will be designed by having the benign violation (McGraw & Warren, 2010) and CATLM (Moreno & Mayer, 2007) theories in mind. Designing IH by the previously mentioned theories meant that the participants' both medium, as well as topic-related mental representations (MRs) had to be violated so as to provide a narration that aids learning in CATLM. Thus, the author had to initially know the participants' MRs.

This study used the mind-map method to gain access to the participants MRs (Ludden et al., 2012). In this study, the bigger circle in the middle of the paper had the words "Brain Cells" written within it, while the smaller circles that surrounded it were left empty. The participants were asked to free associate whatever comes to their mind related to "Brain Cells," thus revealing both medium as well as close MRs related to the chosen instructional topic.

The generated similar MRs of the participants related to the "Brain Cells" topic were grouped into four categories, such as: a) accurate and relevant, b) accurate and irrelevant, c) inaccurate and relevant, and d) inaccurate and irrelevant.

Of the above, inaccurate and relevant/misconceptions were chosen to be benignly violated because they are related to the topic, which would result in adaptive and educationally "appropriate" humor (Suzuki & Heath, 2014; Wanzer et al., 2010). Lastly, the benefits of loading the learners limited working

memory capacities with topic-related adaptive humor may outweigh the risks, since it could result in intrinsic (and not extrinsic) load, which in turn may not harm learning with the seductive detail effect (Park et al., 2015). One example of IH is presented in Figure 2.

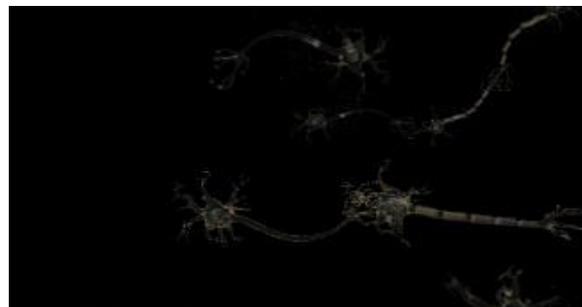


Figure 2. The multimedia stimuli for both nonhumorous and humorous conditions.

Source: Dorambari (2018).

This image is a sample of a cropped screenshot depicting neural activity during the multimedia video presentations. The non-humorous (NH) condition had 21 video sequences, while the humorous (IHCALM) condition had 30 video sequences (to account for both instruction and humor). The IHCALM condition narration during this video sequence was: "Anyway, the 'beauty' of brain cells while they communicate by receiving, processing, and transmitting signals is so enormous, that they say that it has even inspired Leonardo Da Vinci to paint Mona Lisa!" The NH condition narration was: "Brain cells communicate by receiving, processing, and transmitting signals."

Stimuli were computer generated 3D video imageries designed by me that depicted neural activity. This image is a sample of a cropped screenshot depicting neural activity during the multimedia video presentations.

Finally, the IH narration was added a high pitched tone for two reasons. Firstly, it was important (as per CATLM principles) to have shorter video sequences so as not to cognitively overload the participants limited working memory units (Mayer, 2008). Secondly, the high pitched tone of the IH narration helped avoid potential patronizing multimedia voice-related perceptions from the participants.

3.1. Design

Alas, there were two study conditions: 1) The experimental condition that had IH narration in a 3D video depicting brain cells (the IHCALM condition), and 2) the control condition that had just instruction narration about brain cells (without IH) in the same 3D video depicting brain cells (the NH condition). This

produced two forms of multimedia—IHCALM and NH—that were compared for differences with CATLM DVs (see example narration in Figure 2). If IH is a seductive detail, then the mean should significantly be different in favor of the NH condition for all the CATLM DVs.

Since there was only one IV of IH, while there were many CATLM DVs, then a classic experiment with two independent samples was designed in OpenSesame (Mathôt et al., 2012). The participants choose to go through all of cognitive load (n = 5), humor (n = 2), academic emotions (n = 20), motivation (n = 1), learning (n = 2), and metacognition (n = 2) CATLM DVs (N = 30).

3.2. Procedure

The participants were approached in the University cafes or when they were found standing idle, relaxing, inside unused classrooms. After being orally briefed about the study, the interested participants went together with the author to the computer hall, which was a walking distance from the university. The computers had cameras and (sound interfering) headsets installed. Those same headsets now also came to be useful for blocking potential audible interferences from other participants during the experiment.

In addition, the hall had cubicles that separated each other participants physically. The cubicles helped block potential tactile or visual interferences from other participants. Lastly, other than this study's participants, no one else entered during the exclusive reservation hours of the hall booked for this experiment.

When the participants came to the experiment, the computers were mostly ready for use in the experiment. The participants were randomly assigned to a computer, which had OpenSesame with either the IHCALM or the NH conditions. After being seated, the ethical procedures followed.

Upon issuing the briefing sheets and quickly orally re-briefing the participants, OpenSesame presented a consent form to each participant. The consent form in OpenSesame asked the participants whether they understood the information in the previously issued briefing form, which was related to voluntary participation and data confidentiality (participants could choose to click "OK"). OpenSesame would only start the experiment after participants consented to the experiment by clicking the "OK" button on the screen. After providing consent, the webcam began recording the participant's potential mirth.

While the camera recorded, the participants were informed through OpenSesame that they will take part in an initial practice session before engaging with the real experiment. The practice session included the appearance of the four random digits on a black screen. The four random digits were presented to preload the participants' working memory before issuing a multimedia video sequence. The participants were instructed to remember the numbers and press any button to watch the practice session video sequence.

There were five video sequences in total (duration of up to 1 min and 30 s approximately) for the practice session of the experiment for both the IHCALM and the NH conditions. The video sequences were the same, while the narration differed between the stimuli conditions. The IHCALM condition consisted of humor unrelated to the topic, while the narration for the NH multimedia condition consisted of abstract concepts related to brain cell activity. After viewing the corresponding video sequence, the participants were asked to type the four preloaded random digits.

Thus, participants practiced preloading their working memories with random numbers that measured cognitive load. After they typed the random numbers, a page appeared that asked the participants to either repeat the same video sequence or to move on to the next one. Intrinsic cognitive load was measured at this point.

The practice session ended after five such video sequences. The practice session of the experiment lasted up to 2 min (approximately) per participant. After the practice session, OpenSesame informed the participants that the experimental session would now commence.

During the experimental session, the participants were informed that (unlike during the practice session) the content of the video now mattered, as they would be tested on the content at a later stage. They were again preloaded with four random digits, watched a video sequence, typed the preloaded random digits, and clicked whether they wished to repeat the same video sequence again or not (just like in the practice session). The experimental multimedia video sequences were presented to the participants in an order that depended on their randomly assigned condition (experimental/IHCALM or control/NH) groups. After all the video sequences ended, the experiment was briefly stopped to shut down the camera recorder (post-multimedia potential mirth was not important for this research).

Upon stopping the camera recorder, the video recording application Camtasia processed the recording material and

formatted it into a viewable video format for later coding. While Camtasia was working in parallel, the participants were placed back to their assigned experimental conditions. At this point, the experiment moved on to its subsequent sessions.

In the next session, OpenSesame briefed the participants that questions regarding humor, academic emotions, and motivation were to follow. The monotonic humorous scale, which asked the participants to rate the humor in the multimedia content, was issued. Following the humor scale, the twenty PANAS questions related to how the participants felt at the current moment followed. Lastly, the four BAS questions (which measured drive) were issued to participants, and at this point, the experiment moved on to its last session (duration of this session lasted 10–20 min approximately).

During the last session, OpenSesame informed the participants that the tests would now commence. The test consisted of questions ($N = 30$) that measured retention with five multiple-choice answers, followed by a binary question that asked whether the student was certain of their previously chosen answer or not. The participants were informed that they had 2 min to answer each question; the test could technically possibly last only up to 120 min (although most of them took anywhere between 10 and 40 min approximately).

After they completed the test, the participants were thanked and debriefed. At the same time, the Camtasia application (that was working in parallel all along) processed the recordings into videos. The videos together with the OpenSesame data were copied and deleted away from the computers. The entire experiment lasted from 20 to 50 min approximately.

3.3. Participants

The sample was taken during the beginning of the winter semester 2016/2017. Concurrent with the previous CATLM research, the study population consisted of higher education students (aged 18 – 25 approximately). Therefore, the study population consisted of undergraduate students who either had (or were going to have) a class related to the biological basis of behavior in a psychology (or otherwise a brain activity-related) course. As such, the population consists of a convenience sample made up of young students mostly from the psychology department, followed by students from the childcare, nursing, and criminology departments.

Students from the psychology department comprised the largest representation from a single school of study ($N = 48$, IHCALM = 24, NH = 24). Second largest was followed by students from the school of nursing ($N = 25$, IHCALM = 12, NH = 13), while the

third was from the school of criminology ($N = 26$, IHCALM = 15, NH = 11). After dropout, the total number of participants resulted to 96 (IHCALM = 48, NH = 48), where 66 participants were female and 28 were male.

3.4. Measures

3.4.1. Stimuli

The 3D animations about “Brain Cells” depicted sub-topics related to action potentials, myelin versus non-myelin neural cells, and neurotransmitter spatial summation. Since the sub-topics differed in narration (IHCALM had both IH and instruction, while NH had only instruction), then this resulted in a difference in terms of the quantity of the video sequences. Particularly, the IHCALM ($N = 30$) and the NH ($N = 21$) conditions differed by nine additional video sequences for the former. Lastly, there was also a small difference of 36 s in duration between IHCALM (7 min and 19 s) and NH (6 min and 43 s) conditions.

3.4.2. Cognitive load

The cognitive load of multimedia instruction was measured with preload (Brunken et al., 2002). The preload method issues random digits to participants prior to watching stimuli. The digits reliably load the phonological loop of participants’ working memory (Pearson’s $r = 0.80$ – 0.91 [Schuler et al., 2011]). A similar approach was used in the Cocchini et al. (2002) and Kruley et al. (1994) studies.

The preload instrument generated four variables that measured cognitive load (with values of zero and above), such as: 1) error (when students typed the wrong digits), 2) misplaced digits (when students wrote the correct numbers, yet placed them in the wrong places), 3) missing values (when students recalled no [or only partial] numbers from the four random digits), and 4) repeated similar values (when students wrote the correct four digits; however, the written digits were from a previous video sequence preload).

3.4.3. Video sequence repeats

Since the IHCALM condition had both IH and instruction, then this amounted to more video sequences and duration, which also should result in more intrinsic cognitive load. Sequencing is proposed as a solution to multimedia instruction that may have a high intrinsic cognitive load (Mayer, 2008). Therefore, depending on their pace, OpenSesame was programmed to make the option available for participants to review the same multimedia video sequence as many times as they required to ease intrinsic load. This programming generated the video sequence repeats variable, which was a one-item instrument that measured how many times

(i.e., zero and above) the participants had viewed a particular video sequence, which indicated intrinsic load.

3.4.4. Humorous scale

A one-item monotonic scale with four responses measured the participants' self-reported degree of humor that they might have experienced during the experiment. The question was, "How funny did you find the previously viewed multimedia presentation?" to which the participants could answer from "not funny at all," "somewhat funny," "funny," and "funny to a great extent." This variable generated values ranging from zero ("not funny at all") to three ("funny to a great extent").

3.4.5. Mirth

One method to measure mirth is by recording the participants' responses to humor and making use of the facial activation coding system expert (FACS, Ruch et al., 2009). However, since there were imposed limitations in resources, the researcher had to rely on the non-FACS expertise of the former students that volunteered to take the role of independent coders. It was assumed that the student coders would code the participants' mirth naturally just like non-FACS coders did in previous research (Falk & Hill, 1992).

The independent coders were instructed to observe and code participant responses based on findings of Ruch (1993), such as: a) if no mirth is observed, then move on, b) if there is a smile (with or without laughter) shorter than 2/3 of a second, then ignore and move on, c) if there is a smile (with or without laughter) longer than 2/3 of a second, then measure the entire response duration as a "mirth."

3.4.6. Academic emotions

Academic emotions were measured with the positive affect and negative affect scale (PANAS; Watson et al., 1988). Considering that both PAEs and NAEs help with motivation and learning (Pekrun & Stephens, 2012: Ch. 1), as well as considering that both might be incited with IH in the IHCALM condition, then the author decided to use all academic emotions for the first time in CATLM research. The reliability of the PAEs is somewhat higher (Cronbach's $\alpha = 0.89$) than for the NAEs (Cronbach's $\alpha = 0.85$, Watson et al., 1988). The PANAS points could range from 0 to 80.

3.4.7. Behavioral activation system

The behavioral inhibition and behavioral activation scales (Carver & White, 1994) were used to measure the dissonance reduction motivation, which is innately found in any humour (Harmon-Jones and Harmon-Jones, 2007). In addition, both the

mirth reward motivation, as well as learning motivation was also of interest to be measured. Although the instrument was useful to measure motivation from various angles, it also incorporated various scales.

As the purpose of this study was to measure various types of attraction (rather than inhibition), then, only the BAS scale (and not the BIS scale) was used. The BAS element of the instrument was further divided into BAS Reward Responsiveness (Cronbach's $\alpha = 0.73$), BAS Drive (Cronbach's $\alpha = 0.76$), and BAS Fun Seeking (Cronbach's $\alpha = 0.66$). Of the three, the BAS Drive instrument alone was selected for use in this study because of its highest degree of reliability (Carver & White, 1994), as well as to lessen the number overall variables in the study following academic emotions.

3.4.8. Learning

Retention was measured with 30 questions in total, which were related to the multimedia video presentations viewed earlier. Each question had five multiple answers and was presented in a page named "A". The participants could earn a point if they selected the correct answer (among the five) within 2 min. The duration was set to 2 min (rather than unlimited time) to control for the possibility that prolonged duration (rather than the stimuli conditions) may influence the participant's responses.

Therefore, if the participant did not answer (or when the time was up), the question was left unanswered (measured with Missing Answers) and a new question was issued; this produced two variables of Correct Answers and Missing Answers. The total score of these variables could range from 0 to 30 (a similar instrument was used to measure learning in previous CATLM studies, e.g., Um et al., 2011). There was no baseline as prior-knowledge was not measured in this study (unlike in Mayer & Estrella, 2014), and the inter-item reliability measured with Cronbach's alpha was low ($\alpha = 0.46$).

3.4.9. Metacognition

Lastly, metacognition was measured for the first time in CATLM research by how confident the participants were of their previously chosen answer. For each retention question on page "A," a metacognitive question followed and was placed on page "B." The metacognitive question simply asked, "Do you think that your answer to the previous Question A was correct?" – to which the participants had to answer with a binary "Yes" or "No." The interaction between retention on page "A" ("Correct" or "Incorrect") and metacognition on page "B" ("Yes" or "No") produced four outcomes of true positive, true negative, false positive, and false negative (Dienes & Seth, 2010; Fleming & Lau, 2014; Maniscalco & Lau, 2012).

The true positive and negative scores were summed and then divided by the total number of questions to reveal a metacognitive confidence accuracy ratio. The metacognitive confidence accuracy ratio was multiplied with 100 to reveal a metacognitive confidence percent (Dienes & Seth, 2010). This measurement produced two variables of Metacognitive Percent and (since this variable also had a time duration of two minutes) Missing Metacognitive Values that had values from zero to hundred.

The reliability of these metacognitive measures was accounted for with the receiver operator characteristic (ROC) analysis that produced the area under the curve (AUC) value. The NH condition was slightly more reliable in detecting overall sensitivity (AUC = .98, $p < .01$) compared to the IHCALM condition (AUC = .96, $p < .01$).

3.5. Data Analysis

The null-hypothesis significance testing (NHST) was analysed with SPSS, while the Bayesian factor (BF) test was analysed in JASP. If the IHCALM condition is not another type of seductive detail, then the mean of the IHCALM condition should not be significantly different from the NH condition (i.e., $p > .05$). However, since absence of evidence is not evidence of absence, then only the NHST analysis would not suffice. Therefore, to find the evidence that the NH condition is similar to the IHCALM condition the Bayesian analysis was also included, which should also tell how much there is evidence that the two outcomes yield similar results (i.e., “Evidence for the H_0 ” outcome).

Using both analyses could also better inform the reader of the outcomes of the results, which is why it is proposed that the two values should be placed side by side in a table (Quintana & Williams, 2018; Wagenmakers et al., 2018). Since for the NHST analysis what matters by convention is whether the p value is less than .05, while for the BF analysis what matters is a higher ratio number (e.g., >3), then both analyses could reveal four general combinations of outcomes. Those outcomes are presented below in Table 1.

Table 1. Four general outcomes between the null hypothesis testing (NHST) and the Bayesian factor (BF) test

	NHST ($p < .05$)	NHST ($p > .05$)
BF (<3)	Evidence for the H_1	No power, or insensitive instrument
BF (>3)	Humorous or N/A	Evidence for the H_0

The first in the upper left corner on the table could be called the “Evidence for the H_1 ” outcome. This applies when both the

means are significantly different in favour of the alternative hypothesis and there is no evidence for the null hypothesis (NHST: $p < .05$, $BF_{01} = 0 - 1$). This is the best result if the alternative hypothesis is the desired outcome, since both analyses point to the same direction that the results are in favor of the alternative (and not the null) hypothesis.

The second in the lower right could be called the “Evidence for the H_0 ” outcome. This applies when both the means are not significantly different in favor of the alternative hypothesis and there is evidence for the null hypothesis (NHST: $p > .05$, $BF_{01} > 3$). Since the absence of evidence does not mean evidence of absence, then just because NHST results came up as non-significant does not automatically mean that there is evidence for the null hypothesis. Rather than leave it at that, in such cases, the BF analyses could be used as evidence in favor of the null (and not the alternative) hypothesis.

The third in the upper right corner could be called the “insensitive instrument,” or the “no power” outcome. This outcome applies when both the means are not significantly different in favour of the alternative hypothesis and there is no evidence for the null hypothesis neither (NHST: $p > .05$, $BF_{01} = 0 - 1$). Since the outcome is neither, then the reader may conclude that the instrument was not sensitive enough to measure anything outstanding or worthy to see for that DV.

The fourth in the lower left is mentioned here largely to complement the combinations of the outcomes mentioned above. This last one could be called as the “N/A” or “humorous” outcome, and it would apply when both the means are significantly different in favor of the alternative hypothesis and there is evidence for the null hypothesis just as well (NHST: $p < .05$, $BF_{01} > 3$). This outcome may leave the researcher initially in a dissonant state as to why did these results occur, only to realize that it was probably a miscalculation, whereupon the previous dissonance tension is then released in exhilaration. As entertaining as this outcome may be, it was not found in this, or any other study that I have read thus far.

4. Results

All participants were randomly assigned to either the IHCALM or the NH condition and their data were gathered with the OpenSesame program. There are no side or adverse effects to report. Lastly, since this was a preliminary study about IH in CATLM, then the two conditions were just analyzed for similarities and differences based on Outcomes mentioned in Table 1.

The **Table 2** below presents descriptive statistics and parametric tests for the CATLM dependent variables. Standard deviations

are shown below the means in brackets. The distribution was normal for most dependent variables (K-S, $p > .05$).

Table 2. The mean, standard deviation, F-values, Bayesian Factors, and effect sizes for humor, cognitive load, academic emotions, motivation, learning, and metacognition variables.

Dependent Variables	NH		IHCALM		
	M (SD)	M (SD)	F	BF^{01}_a	Cohen's <i>d</i>
Mirth	2.92 (6.26)	8.63 (17.03)	4.75* _b	0.58	0.45
Humorous scale	0.17 (0.6)	1.52 (0.95)	70.54*** _b	0.01	1.51
Cognitive load error	13.85 (7.01)	18.33 (17.6)	1.61	2.29	0.21
Cognitive load misplaced digits	4.06 (4.32)	6.54 (6.37)	4.97*	0.53	0.4
Cognitive load missing values	3.69 (9.39)	3.81 (6.04)	0.01	4.65	0.01
Cognitive load similar repeating values	0.52 (1.94)	0.94 (2.75)	0.74	3.36	0.15
Video sequence repeats	1.9 (4.22)	1.65 (3.8)	0.09	4.47	0.05
Interest	2.15 (1.5)	2.44 (1.52)	0.90	3.13	0.16
Excitement	3.42 (1.33)	3.4 (1.38)	0.01	4.65	0.01
Strong	3 (1.27)	2.71 (1.24)	1.30	2.63	0.19
Enthusiasm	2.83 (1.31)	2.81 (1.42)	0.01	4.65	0.01
Proud	2.6 (1.28)	2.67 (1.24)	0.06	4.54	0.05
Alert	3.38 (1.1)	3.6 (1.27)	0.89	3.14	0.15
Inspired	3.23 (1.33)	3.06 (1.52)	0.33	4	0.1
Determined	2.83 (1.21)	2.85 (1.2)	0.01	4.64	0.01
Attentive	3.73 (1.04)	3.6 (1.33)	0.26	4.15	0.09
Active	3.27 (1.28)	3.46 (1.22)	0.54	3.67	0.12

Ashamed	1.42 (0.74)	1.71 (1.13)	2.24 ^b	1.73	0.27
Nervous	1.46 (0.87)	1.79 (1.44)	1.87 ^b	2	0.25
Distress	1.75 (1.23)	2.15 (1.3)	2.34	1.66	0.26
Upset	1.9 (1.15)	1.81 (1.21)	0.17	4.42	0.06
Guilty	1.25 (0.86)	1.52 (0.92)	2.21	1.76	0.25
Scared	1.4 (0.82)	1.46 (0.94)	0.12	4.42	0.06
Hostile	1.48 (1.03)	1.65 (1.14)	0.56	3.63	0.13
Irritated	1.52 (1.01)	1.88 (1.12)	2.64	1.46	0.28
Jittery	2.02 (1)	2.23 (1.13)	0.91	3.11	0.16
Afraid	1.44 (0.92)	1.33 (0.86)	0.33	4.03	0.1
BAS	0.07 (2.54)	0.89 (3.5)	2.28	1.32	0.31
Correct answers	12.08 (3.2)	11.35 (3.35)	1.19	2.76	0.18
Missing answers	0.25 (0.56)	0.27 (0.94)	0.02	4.62	0.02
Metacognitive percent	51.73 (11.4)	44.44 (12.91)	8.60 ^{**}	0.11	0.5
Metacognitive missing values	0.23 (0.69)	0.52 (0.92)	3.07 ^b	1.21	0.31

a BF01 Bayesian factor analysis results. b The variances were nonhomogeneous ($p < .05$), thereby Robust Welch analysis of variance was used for these dependent variables.

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

Since the humour variables did not meet the parametric assumptions, the robust Welch analysis of variance (ANOVA) was used. The analysis revealed that the variable humourous scale as well as mirth favoured the “Evidence for the H1” outcome. This was so because firstly, the means for both humour variables were significantly different and in favor of the IHCALM condition. Secondly, the Bayesian Factor (BF) analysis provides near-zero evidence that the conditions are similar (see Tables 1 & 2).

Following the manipulation check, the first hypothesis to be tested was whether there was any difference in cognitive load. As

there was only one IV of the stimulus conditions and there were 5 DVs of cognitive load that met the parametric assumptions (see Table 2), then the MANOVA was used. However, the multivariate statistic revealed no significant differences in the cognitive load variables, $F(5, 90) = 1.178, p = .326$.

In the follow up ANOVA and BF analysis, it was revealed that the data favoured the “Evidence for the H0” outcome. To reiterate, the “Evidence for the H0” outcome applies when the author both fails to reject the null hypothesis and BF analysis indicates similarity between conditions with a high ratio number, which was the case for most of the cognitive load DVs. Thus,

participants did not experience higher cognitive load, despite larger narration in the IHCALM condition.

The next hypothesis to be tested was for academic emotions. Initially, all academic emotions ($N = 20$) were halved into PAE ($n = 10$) and NAE ($n = 10$) subgroups. Preliminary analysis revealed the variables ashamed and nervous did not meet the parametric tests (Table 2). Thereby, the PAE and most NAE subgroups were analyzed with MANOVA; ashamed and nervous variables were analyzed with the robust Welch ANOVA. Finally, all academic emotions also underwent BF analysis.

The analysis again favoured the “Evidence for the H_0 ” outcome for the majority of academic emotions. The MANOVA was neither significant for PAE, $F(10, 85) = 0.623$, $p = .790$, nor for NAE $F(8, 87) = 0.930$, $p = .496$ subgroups (see Table 2 for ashamed and nervous NAEs analyzed in Welch ANOVA). Based on the data from the follow up ANOVA and BF analysis, the author both fails to reject the null hypothesis and finds evidence that the conditions are similar for academic emotions also, indicating that the data favor the “Evidence for the H_0 ” outcome. The next hypothesis to be tested was about motivation. However, the ANOVA again favoured the “Evidence for the H_0 ” outcome (see Table 2). With this outcome, the author both fails to reject the null hypothesis and finds evidence that the conditions are similar for motivation.

The remaining critical (apart from the manipulation) variables that were going to be tested relate to learning. As the correct and missing answers variables met the parametric assumptions, then ANOVA and BF analysis was carried out. The analysis was again in favour of the “Evidence for the H_0 ” outcome for both correct answers and missing answers DVs (see Table 2). Again, with this outcome the author both fails to reject the null hypothesis and finds evidence that the conditions are similar for learning as well

The last remaining variable to be tested was the metacognitive percent variable, which measured the certainty of the participants’ previously chosen answer. As the variable metacognitive percent met the requirements of the parametric tests, it was analyzed with ANOVA. Since the metacognitive missing values had nonhomogeneous variances, then the robust Welch ANOVA was used instead (see Table 2). Lastly, and similarly to the previous analysis, both DVs were also analyzed with BF to check for similarities between conditions.

This time, the one-way ANOVA and BF analysis for metacognitive percent favoured the “Evidence for the H_1 ” outcome for a change, but (ironically, and contrary to expectations) the outcome was found to be in favour for the NH

condition instead. Although the missing metacognitive values variable was near significantly higher for the IHCALM condition ($p = .08$), it was conventionally non-significant. Thereby, the author both rejects the null hypothesis and finds evidence that the conditions are not similar for metacognition; the data favour “Evidence for the H_1 ” where metacognition values were both significantly different and higher in the NH condition, compared to the IHCALM condition.

5. Discussion

Since the data point to the “Evidence for H_1 ” outcome for the humorous scale and mirth variables, then this meant that the manipulation was successful and that students did find the IHCALM condition to be more humorous than the NH condition. Thus, the findings in this study demonstrated that benignly violating the students’ misconceptions resulted in IH, which caused significantly different multimedia experiences with more exhilaration for the IHCALM condition.

Despite IH, the data point to the “Evidence for the H_0 ” outcome for academic emotions, motivation, as well as most learning variables. If IH harmed learning via seductive detail, then the “Evidence for the H_1 ” would have resulted in favor of the NH condition. Since this is not the case, then the data indicate that IH in the IHCALM condition does not harm learning and is not a seductive detail type for most measured CATLM DVs in this research, except metacognition.

A contrary to expectation outcome was found for the other learning variable, metacognitive percent. The data analysis favoured the “Evidence for the H_1 ” outcome in favour of the NH condition, despite the condition only coding pure instruction without humour (i.e., lesser sources to gather data from LTM during retrieval). However, the participants under the IHCALM condition (17 out of 48, 35.42%) did not report their certainty of the previously chosen answers compared to the NH condition (7 out of 48, 14.58%) to a considerable degree. Since more data in the metacognitive missing values variable means fewer data in the metacognitive percent variable, then the lack of enough data for the metacognitive percent variable may have shadowed the true difference between the stimuli conditions.

5.1. Empirical Contributions

The data presented in this paper challenge previous findings on intrinsic cognitive load, the effect of academic emotion on motivation and their effect in turn on learning outcomes. Perhaps a future study with a larger sample may shed more light into this matter.

5.2. Practical and Theoretical Implications

Even so, this study demonstrated that the mind-map method could be used to design a class specific IH and present a CATLM video that does not harm learning. As such, the study also presents a new theory of learning with IHCALM, which is found not to be another seductive detail effect. As such, IHCALM can be used in education.

5.3. Limitations and future directions

A future study may do better by increasing the sensitivity and specificity of the instrument that measured academic emotions. Since the NHST was always non-significant for these variables, while the BF analysis ranged from 1.46 – 4.65, then the outcome also ranged from “insensitive instrument” (or “no power”) to “Evidence for H₀” outcomes per individual academic emotion variables (Tables 1 and 2). Since there is evidence that the instrument was insensitive for some academic emotions in this study, then a more sensitive and specific measurement may be sought in future studies.

Individual differences that should be attended in future studies are cognitive covariates such as prior knowledge, working memory capacity (Lusk et al., 2009), or even humor predisposition. Without accounting for the covariates, it could be stated that it may be the contribution of the covariates that have resulted in the “Evidence for the H₁” outcome in favour of the IHCALM condition during the manipulation check. For example, the participants in the IHCALM condition may have had personalities with order needs (Ruch & Hehl, 1993), were tough/tender-minded, or were the sensation-seeking type (Ruch, 1988), which renders them more sensitive to humor. Thereby, it could be stated that in such an instance that the “Evidence for the H₁” outcome in favour of the IHCALM condition may have been due to the participants and not the IV (hence, covariates should be controlled in future research).

6. Conclusion

Based on the data gathered above, some conclusions can now be drawn. Since the majority of CATLM DVs were both not significant and the Bayesian factor analysis ratio was around 3 approximately, then this demonstrates that the results are similar between the IHCALM and NH conditions (i.e., “Evidence for the H₀” outcome). Thereby, this study indicates that IHCALM is not harmful in CATLM because otherwise most of the DVs would be both statistically significant and in favor of the NH condition (i.e., the “Evidence for H₁” outcome).

Thus, it can be concluded that IHCALM is not another type of seductive detail that harms learning, unlike emotionally appealing adjuncts, spectacular videos, or soothing background music (Harp

& Mayer, 1998; Park et al., 2015). When an IHCALM multimedia presentation is designed to target the average misconceptions of participants, then the participants were rewarded with humorous mirth, which helped students learn much like in the NH condition despite having more words, video duration, and more video sequences. It can be further added that IHCALM should be used in multimedia presentations because both it is not another type of a seductive detail effect and (being similar with the NH condition) makes the former a priori more advantageous because of humour, which is linked with added psychological, physiological, and social benefits (Morrison & Quest, 2012; Wilkins & Eisenbraun, 2009).

References

1. Brunken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual task methodology. *Experimental Psychology*, 49(2), 109-119. <https://doi.org/10.1027/1618-3169.49.2.109>
2. Carver, C. S., & White, T. L. (1994). Behavioural inhibition, behavioural activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology*, 67(2), 319-333. <https://doi.org/10.1037/0022-3514.67.2.319>
3. Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, 30(7), 1086-1095. <https://doi.org/10.3758/BF03194326>
4. Dienes, Z., & Seth, A. (2010). Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Consciousness and Cognition*, 19(2), 674-681. <https://doi.org/10.1016/j.concog.2009.09.009>
5. Falk, D. R., & Hill, C. E. (1992). Counselor interventions preceding client laughter in brief therapy. *Journal of Counseling Psychology*, 39(1), 39-45. <https://doi.org/10.1037/0022-0167.39.1.39>
6. Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00443>
7. Franklin, R. G., Jr., & Adams, R. B., Jr. (2011). The reward of a good joke: neural correlates of viewing dynamic displays of stand-up comedy. *Cognitive, Affective, & Behavioral Neuroscience*, 11(4), 508-515. <https://doi.org/10.3758/s13415-011-0049-7>

8. Greengross, G., Martin, R. A., & Miller, G. (2012). Personality traits, intelligence, humour styles, and humour production ability of professional stand-up comedians compared to college students. *Psychology of Aesthetics, Creativity, and the Arts*, 6(1), 74 - 82. <https://doi.org/10.1037/a0025774>
9. Hackathorn, J., Garczynski, A. M., Blankmeyer, K., Tennial, R. D., & Solomon, E. D. (2012). All kidding aside: Humour increases learning at knowledge and comprehension levels. *Journal of the Scholarship of Teaching and Learning*, 11(4), 116-123. ISSN: ISSN-1527-9316
10. Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, 38(1), 7-16. <https://doi.org/10.1024/0044-3514.38.1.7>
11. Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414. <https://doi.org/10.1037/0022-0663.90.3.414>
12. Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behaviour*, 44, 81-95. <https://doi.org/10.1016/j.chb.2014.11.009>
13. Kruley, P., Sciamia, S. C., & Glenberg, A. M. (1994). On-line processing of textual illustrations in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition*, 22(3), 261-272. <https://doi.org/10.3758/BF03200853>
14. Lovorn, M. G. (2008). Humour in the home and in the classroom: The benefits of laughing while we learn. *Journal of Education and Human Development*, 2(1). ISSN: 1934-7200
15. Ludden, G. D., Kudrowitz, B. M., Schifferstein, H. N., & Hekkert, P. (2012). Surprise and humour in product design. *Humour*, 25(3), 285-309. Humour in the home and in the classroom: The benefits of laughing while we learn. <https://doi.org/10.1515/humour-2012-0015>
16. Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology*, 40(4), 636-651. <https://doi.org/10.1111/j.1467-8535.2008.00848.x>
17. Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422-430. <https://doi.org/10.1016/j.concog.2011.09.021>
18. Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behaviour Research Methods*, 44(2), 314-324. <https://doi.org/10.3758/s13428-011-0168-7>
19. Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *American Psychologist*, 63(8), 760-769. <https://doi.org/10.1037/0003-066X.63.8.760>
20. Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction*, 33, 12-18. <https://doi.org/10.1016/j.learninstruc.2014.02.004>
21. McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological science*, 21(8), 1141-1149. <https://doi.org/10.1177/0956797610376073>
22. Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309-326. <https://doi.org/10.1007/s10648-007-9047-2>
23. Morrison, M. K., & Quest, H. (2012). The top ten reasons why humour is Fundamental to education. *Creating an Appropriate 21 st Century Education*, 48-50.
24. Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behaviour*, 44, 267-278. <https://doi.org/10.1016/j.chb.2014.10.061>
25. Pekrun, R., & Stephens, E. J. (2012). Academic emotions. *Academic Learning & Achievement*, 3-31. <https://doi.org/10.1037/13274-001>
26. Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *Bio Medical Center Psychiatry*, 18(1), 178 - 186. <https://doi.org/10.1186/s12888-018-1761-4>
27. Ruch, W. (1988). Sensation seeking and the enjoyment of structure and content of humour: Stability of findings across four samples. *Personality and Individual Differences*, 9(5), 861-871. [https://doi.org/10.1016/0191-8869\(88\)90004-9](https://doi.org/10.1016/0191-8869(88)90004-9)

28. Ruch, W. (1993). Exhilaration and humour. *Handbook of Emotions*, 1, 605-616. <https://doi.org/10.5167/uzh-77841>
29. Ruch, W., & Hehl, F. J. (1993). Humour appreciation and needs: Evidence from questionnaire, self-, and peer-rating data. *Personality and Individual Differences*, 15(4), 433-445. [https://doi.org/10.1016/0191-8869\(93\)90071-A](https://doi.org/10.1016/0191-8869(93)90071-A)
30. Ruch, W., Bänninger-Huber, E., & Peham, D. (2009). Unresolved issues in research on humour and laughter: The need for FACS-studies (pp. 42-46). Innsbruck University Press.
31. Schuler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review*, 23, 389-411. <https://doi.org/10.1007/s10648-011-9168-5>
32. Suzuki, H., & Heath, L. (2014). Impacts of humour and relevance on the remembering of lecture details. *Humour*, 27(1), 87-101. <https://doi.org/10.1515/humour-2013-0051>
33. Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2011). Emotional design in multimedia. *Journal of Educational Psychology*, 104(2), 485-498. <https://doi.org/10.1016/j.jchb.2014.11.009>
34. Vrticka, P., Black, J. M., & Reiss, A. L. (2013). The neural basis of humour processing. *Nature Reviews Neuroscience*, 14(12), 860-868. <https://doi.org/10.1038/nrn3566>
35. Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q.F., Šmíra, M., Epskamp, S. and Matzke, D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58-76. <https://doi.org/10.3758/s13423-017-1323-7>
36. Wanzer, M. B., Frymier, A. B., & Irwin, J. (2010). An explanation of the relationship between instructor humour and student learning: Instructional humour processing theory. *Communication Education*, 59(1), 1-18. <https://doi.org/10.1080/03634520903367238>
37. Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54(6), 1063. <https://doi.org/10.1037/0022-3514.54.6.1063>
38. Wilkins, J., & Eisenbraun, A. J. (2009). Humour theories and the physiological benefits of laughter. *Holistic Nursing Practice*, 23(6), 349-354. <https://doi.org/10.1097/HNP.0b013e3181bf37ad>