International Journal of Education and Practice
2022 Vol. 10, No. 2, pp. 182-203.
ISSN(e): 2310-3868
ISSN(p): 2311-6897
DOI: 10.18488/61.v10i2.3006

INVESTIGATING THE EFFECTS OF INSTRUCTIONAL HUMOR IN MULTIMEDIA LEARNING WHEN HUMOR PRE-DISPOSITION, PRIOR-KNOWLEDGE, AND WORKING MEMORY ARE CONTROLLED

Diedon Dorambari
Heimerer College, Prishtina, Dardania, Kosovo.
Email: d_diedon@hotmail.com

Article History
Received: 4 March 2022
Revised: 22 April 2022
Accepted: 12 May 2022
Published: 26 May 2022

Keywords
BFDA
CATLM
DKE
IHCALM
Metacognitive calibration
Seductive detail.

ABSTRACT

This study examined whether instructional humor (IH) was not just another type of seductive detail when covariates such as humor pre-disposition, prior-knowledge, and working memory capacity were controlled. Participants were students (N = 228) from universities who were randomly assigned two stimuli conditions in the classic experimental design. The data analysis involved a MANCOVA in SPSS and ANCOVA in R-WRS2 package (for DVs with non-homogenous variances) to control the covariates. The data from both null hypothesis significance testing and Bayesian factor design analysis showed that the data were in favor of outcomes which demonstrated that although IHCALM was funnier (p < .01) it was not another type of a seductive detail that harmed learning (p > 0.05). It was less interesting (p < .01), yet made the participants more aware of what they did not learn (p < 0.01). The practical and theoretical implications of teaching with IHCALM were also discussed.

Contribution/Originality: The originality in IHCALM involves identifying student' misconceptions related to the topic via a mind-mapping method. This study generated instructional humor by benignly violating misconceptions and providing narrated instruction. This CATLM study is the first that measured all of the students' academic emotions and (especially) their metacognition.

1. INTRODUCTION

In the Cognitive Affective Theory of Learning with Media (CATLM) (Moreno & Mayer, 2007), findings have demonstrated that emotionally appealing shapes, color, and decorative pictures were not found to be types of the seductive detail effect (Park, Flowerday, & Brünken, 2015) but instead were new elements that fostered learning; classified as multimedia with emotional design (Heidig, Müller, & Reichelt, 2015; Mayer. & Estrella, 2014; Plass, Heidig, Hayward, Homer, & Um, 2014; Schneider, Nebel, & Rey, 2016; Schneider, Nebel, Beege, & Rey, 2018; Um, Plass, Hayward, & Homer, 2012). In a recent CATLM study (Dorambari, 2022) instructional humor (IH) was used as an IV and compared with the non-humorous (NH) condition. The study found that IH was also not a seductive detail, similar with the NH condition (thus did not harm learning) and as such could be used in education. The study offered a new educational approach in CATLM named instructional humor and cognitive affective learning with multimedia (IHCALM). However, the study did not offer any control to covariates, such as prior-knowledge, humor pre-disposition and working memory capacity (WMC). As such, the role of above-mentioned covariates on

IHCALM was not known, which may place the application of IHCALM in education into question. Depending on the effects of covariates on IHCALM, it may turn out that this teaching approach would not reliably result in a non-seductive detail effect. It may also harm learning and thus should not be used in education, which might spare students from the joys of learning with humor in CATLM presentations. Thereby, the purpose of this study aimed to re-demonstrate that IHCALM was not seductive in two ways: direct replication and controlling the above-mentioned covariates. In this study, instructional humor (IH) was used as an IV and compared with the non-humorous (NH) condition. All other variables were identified as CVs and DVs in this study. Table 1 presents a list of all these mentioned variables, which is then followed by a theoretical framework in Figure 1.

Table 1. The independent, covariate and dependent variables in the current IHCALM study.

IHCALM Variables	
Independent Variable	
Instructional Humor	
	Covariate variables
1.	Humor-Predisposition
2.	Prior-Knowledge
3.	Digit Span
4.	Arrow Span
	Dependent Variables
1.	Humorous Scale
2.	Mirth
3.	Cognitive Load Error
4.	Cognitive Load Misplaced Digits
5.	Cognitive Load Missing Values
6.	Cognitive Load Repeated Similar Values
7.	Repeats
8.	Afraid
9.	Distress
10.	Upset
11.	Scared
12.	Ashamed
13.	Hostile
14.	Nervous
15.	Guilty
16.	Jittery
17.	Irritate
18.	Strong
19.	Interest
20.	Proud
21.	Alert
22.	Inspire
23.	Determined
24.	Excited
25.	Enthusiastic
26.	Attentive
27.	Active
28.	BAS
29.	Correct Answers
30.	Missing Answers
31.	Metacognitive Percent
32.	Metacognitive Missing Values
33.	TP
34.	TN
35.	FP
36.	FN

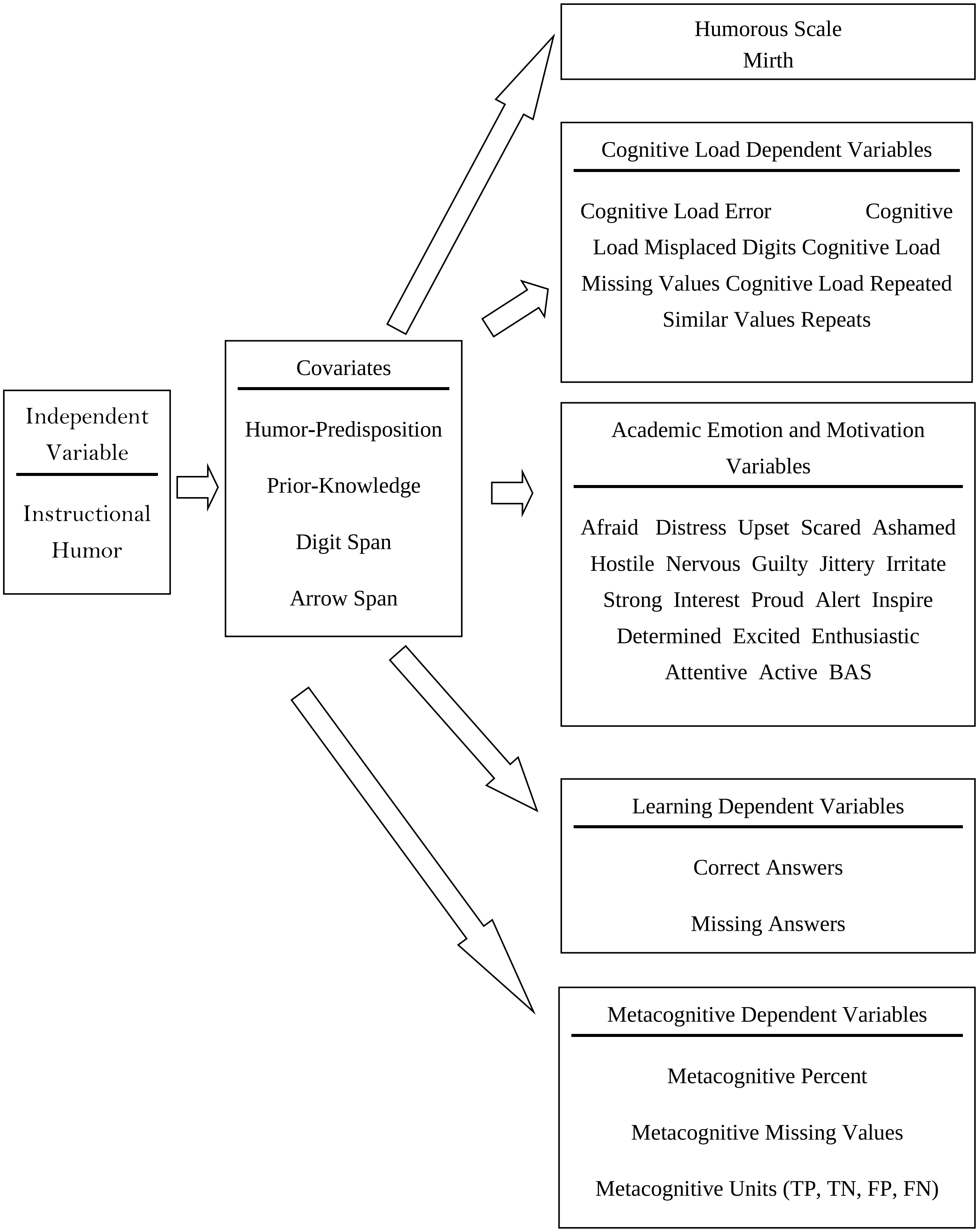

Figure 1. Theoretical framework of all variables in the current IHCALM study.

2. LITERATURE REVIEW

2.1. Dependent Variables

Stimuli: The participants watched a multimedia presentation about "Brain Cells," that were animated in 3D by the author. The brain cell imageries were associated with instructions about action potentials, myelin versus non-myelin neural cells, and neurotransmitter spatial summation. Since the IHCALM condition had both IH and instruction while NH had only instruction, the former had 31 video sequences (duration 7 min and 19 s), while the latter had 21 (duration 6 min and 43 s).

Cognitive Load: Preload is one method to measure cognitive load (Brunken, Steinbacher, Plass, & Leutner, 2002). Random digits were issued to participants prior to watching a video sequence, which loaded their working memory (Schuler, Scheiter, & van Genuchten, 2011). A similar approach was used in the Cocchini, Logie, Della Sala, MacPherson, and Baddeley (2002) and Kruley, Sciama, and Glenberg (1994) studies. The outcome of the preload instrument resulted in four variables, such as: 1) Cognitive load error, 2) cognitive load misplaced digits, 3) cognitive load missing values, and 4) repeated similar values.

Video Sequence Repeats: Video sequencing is a proposed method of lowering intrinsic load (Mayer, 2008). If a video sequence was high in intrinsic cognitive load, OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) was programmed to allow participants to review it. Thereby in this study, the more that a participant reviewed a video sequence, the more this indicated that the video sequence had too much intrinsic load.

Humorous scale: The humorous scale DV was a one-item monotonic scale that measured participant' self-reports of experienced humor. The item asked the participants "How funny did you find the previously viewed multimedia presentation?" to which they could answer from "not funny at all," (0 points) "somewhat funny," (1 point) "funny,"(2 points) and "funny to a great extent"(3 points). The item was issued after the participants had watched all the video sequences.

Mirth: All students volunteered to take the role of independent coders for mirth. It was assumed that the student coders would code the participants' mirth naturally, just like in a previous research (Falk & Hill, 1992). The independent coders were instructed to observe and code participant' mirth based on duration, such as: a) If no mirth is observed, then move on, b) if less than 2/3 of a second, then ignore and move on, c) if more than 2/3 of a second, then measure the entire response duration as a "mirth." These estimates were based on previous humor measuring studies (Ruch, 1993).

Academic emotions. The influence of academic emotions on learning were introduced by studies of Pekrun and Stephens (2012). It was proposed that both positive and negative academic emotions helped with learning. Since instructional humor in the IHCALM condition could incite both negative and positive academic emotions, then it was decided to measure all academic emotions for the first time in CATLM research. This was done by the positive affect and negative affect scale (PANAS) (Watson, Clark, & Tellegen, 1988).

Behavioral activation system. Although academic emotions incite motivation to learn (Pekrun & Stephens, 2012), they are not the only type of motivation that was expected in IHCALM research. Since humor incites dissonance as well as mirth incites reward, both dissonance reduction and reward motivation were expected. Thus to measure motivation, the drive sub-instrument of the behavioral activation scales (Carver & White, 1994) was used, just like in previous research (Harmon-Jones & Harmon-Jones, 2007).

Learning. Retention was measured with 30 questions that had 5 multiple answers each; they were related to the multimedia video sequences. Participants had 2 minutes to select the correct answer, else the next question would follow. This produced two variables of Correct Answers and Missing Answers.

Metacognition. Lastly, for every retention question a metacognitive question followed, which asked "Do you think that your answer to the previous Question A was correct?" The participant could answer with a binary "Yes" or "No." The interaction between Retention and the binary answer produced four outcomes of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which are metacognition measuring units (Dienes &

Seth, 2010; Fleming & Lau, 2014; Maniscalco & Lau, 2012). The Metacognitive Percent variable was calculated by summing TP and TN values, dividing by 30 (total number of questions), and multiplying by 100. If participants did not provide the binary answer, the Metacognitive Missing Values variable was produced (since this variable also had a time duration of two minutes).

2.2 Covariate Instruments

Prior-knowledge. Basic 8 questions were issued to participants via OpenSesame. The questions were related to cells, access to science and scientific instruments, and biology. These same questions were used in CATLM research in the Mayer. and Estrella (2014) study to measure prior-knowledge. For example, a statement was issued "I can name most of the cell's organelles from memory," to which the student participant was asked "How much do you agree on the issued statement?" The participants had 5 options to select from the Likert type scale. The scale ranged from 1) Strongly Disagree (-2 points), 2) Disagree (-1 point), 3) Maybe (0 points), 4) Agree (1 point), and 5) Strongly Agree (2 points).

Humor pre-disposition. Ten jokes were selected from a site. The participants were instructed to evaluate the jokes by selecting from a non-monotonic scale that had items of "Not funny at all" (0 points), "Somewhat funny" (1 point), "Funny" (2 points), and "Very funny" (3 points). The degree that participants found the jokes to be funny indicated their humor pre-disposition.

WMC. The default Cog-Tasks (Stone & Towse, 2015) was used to measure the visuospatial sketchpad and digit-span of WMC (Baddeley, 2012). The Cog-Tasks arrow-span application was used to measure the former, while the digit-span application the latter. Arrow-span was found to be slightly more complicated and mostly disliked by the participants, due to its higher cognitive load. Thereby, the arrow-span application was presented first, prior to participants being too tired for the following digit-span test. The arrow-span application briefly presented arrows on the screen; an arrow had either a small or a large shape that pointed to one of the eight directions (left, right, up, down, diagonal upper left, diagonal lower left, diagonal upper right, and diagonal lower right). The arrows were presented sequentially. There were only two arrows for the first three trials presented sequentially. However, the number of sequentially issued arrows increased all the way up to 7 in the last three trials. The total number of trials was 18. Only when the participants correctly selected all the presented sequential arrows in their exact order in one trial they would receive a point. Since participants had to load their visuospatial sketchpad working memories with size, direction, and an increasing number of arrows in a trial to only score one point (if successful), this demanded too much cognitive load from participants (hence disliked). Thus, none scored more than 7 points with this instrument. The application Cog-Tasks default digit-span was used to measure the phonological loop of working memory. Participants were presented briefly with sequential numbers that had either one or two digits in a trial. After the sequential presentation of all numbers, the participants were asked to type them in their exact order. Initially, only two sets of numbers were presented in five trials. However, the numbers of each set of numbers were increased per every five trials. There were 7 sets of numbers in the last five trials. The total number of trials was 81. The participants scored a point per each correctly typed number in its trial. In other words, it was the correct scoring of an individual number in its order within a trial that scored the participants a point (and not the whole trial itself like in the arrow-span test above). This instrument had less cognitive load and thus was more liked by the participants.

3. METHOD

3.1. Participants

The total number of participants after screening was 228 (Male = 66, Female = 162). All participants were either undertaking the course of Psychology or had studied it at least once. They were mostly young students (M = 20.93, SD = 3.9) from the 1st or 2nd academic year.

3.2. Procedure

The procedures and settings in this study were divided into two parts. In the first part, students were seated in front of computers and OpenSesame (Mathôt et al., 2012). Initially only the demographic data, such as age, academic year, gender, name, and surname was gathered, followed by humor pre-disposition and prior knowledge covariate measurements. The remaining covariate WMC was measured with digit-span and arrow-span that were issued via a different application named Cog-Tasks (Stone & Towse, 2015). Since the participants were already randomly assigned to a computer, which had OpenSesame with either the IHCALM or the NH conditions ready to be used, the second part of the procedure followed. The participants initially practiced in OpenSesame before starting the real experiment. During practice, four random digits were issued, prior to a video sequence. The participants were instructed to remember the numbers and press any button to watch the practice session video sequence, and later type the numbers. After five such practice video sequences, the experiment started. During the experiment, the participants were now informed that the instruction in the content of the video sequence mattered, since a test would follow later. After all experimental video sequences ended, OpenSesame briefed the participants that questions regarding humor, academic emotions, and motivation were to follow. After completing those questionnaires, OpenSesame informed the participants that learning and their metacognitive certainty were going to be measured. Lastly after the test was completed, the participants were thanked and debriefed.

3.3. Design and Analysis

This study followed a classic experimental research design and analyzed the data with both null hypothesis significance testing (NHST) and Bayesian factor design analysis (BFDA) (Quintana & Williams, 2018; Wagenmakers et al., 2018). The covariates were controlled in SPSS for DVs with homogenous variances, while the statistical package R – WRS2 robust ANCOVA was used for DVs with non-homogenous variances.

4. RESULTS

4.1. Preliminary Analysis

Only humor pre-disposition, digit-span, behavior activation system (BAS) (in the IHCALM condition alone), correct answers, and metacognitive percent (in the NH condition alone), were found to be normal; all other covariates and DVs were not normal ($p < 0.01$). Since most of the variables remained significantly non-normal even after transformations, it was decided to use actual data and robust methods of analysis instead. Thereby, the DVs that had non-homogenous variances and regression slopes were analyzed with robust ANOVA in SPSS and later with ANCOVA in the robust R-WRS2 package; both were followed by BFDA in JASP The rest of the DVs with homogenous variances and regression slopes were analyzed with MANOVA and MANCOVA in SPSS and later in JASP for BFDA. The latest way to analyze the covariate influence on DVs that have violated the homogeneity of variance and regression slope assumptions was with the robust ANCOVA in R-WRS2 package (Field & Wilcox, 2017; Mair & Wilcox, 2018). A limitation of this package was that an analysis could only be done for one IV with 2 groups (IHCALM and NH), one DV, and only one covariate at a time. Thereby, all covariates could not be controlled at once in a single analysis with this approach on this study. Thus, each DV was calculated against the IV and one covariate at a time; the covariates' influence on the IV was analyzed separately from the potential joint influences of other covariates combined. This package calculated several trimmed means in robust and representative locations via a non-parametric regression fit across the spread data on non-homogenous variances. As such, the difference with this approach was that on a single analysis of one IV and one covariate, there is more than one F-value shown per each few trimmed means on their most representative locations of the same DV distribution. Thereby, there were also just as many probability values on the same DV for each of its F-values, rather than just one like it is the case with SPSS analysis.

Table 2. The mean, standard deviation, robust ANOVA and ANCOVA F-values, Bayesian factors, and effect sizes for humorous scale, mirth, repeats, and afraid variables that had non-homogenous variances and regression slopes.

Dependent Variables	NH M (SD)	IHCALM M (SD)	$F_{\text{ROBUST ANOVA}}$ (BF_{01}^{x})	$F_{\text{ROBUST ANCOVA}}$ (BF_{01}^{x})	Cohen's d
Cognitive Load Error	8.48	12.23	8.09***	0.33a	-0.37
	(6.61)	(12.56)	(0.16)	2.17a*	
				1.93a	
				2.09a*	
				0.76a	
				(1.14a)	
				0.24b	
				0.8b	
				1.79b	
				2.32b*	
				2.4b*	
				(0.67b)	
				2.00c	
				2.26c*	
				0.78c	
				1.31c	
				1.5c	
				(0.0c)	
				1.8d	
				0.63d	
				1.04d	
				0.25d	
				(0.01d)	
Humorous Scale	0.17	1.43	149.77***	2.58a*	-1.64
	(0.59)	(0.91)	(0.01)	8.50a***	
				16.13a***	
				14.95a***	
				6.70a***	
				(0.01a)	
				11.70b***	
				17.40b***	
				9.97b***	
				15.26b***	
				3.09b***	
				(0.01b)	
				2.45c	
				4.78c***	
				8.93c***	
				16.62c***	
				8.99c***	
				(0.01c)	
				7.80d***	
				16.28d***	
				12.01d***	
				6.69d***	
				(0.01d)	
Mirth	3.12	18.69	19.85***	1.32a	-0.59
	(6.07)	(37.04)	(0.01)	3.81a***	
				4.08a***	
				3.30***	
				2.22*	
				(0.01a)	
				2.62b**	
				3.83b***	
				4.38b***	
				4.68b***	

				2.78b**	
				(0.01b)	
				2.15c*	
				3.03c***	
				4.09c***	
				4.34c***	
				1.33c	
				(0.01c)	
				2.52d**	
				3.04d***	
				4.47d***	
				2.74d**	
				(0.01d)	
Afraid	1.54	1.59	0.17	0.34a	-0.05
	(0.95)	(1.02)	(6.39)	0.15a	
				0.13a	
				0.06a	
				0.52a	
				(1.66a)	
				1.07b	
				0.46b	
				0.71b	
				0.07b	
				0.33b	
				(30.44b)	
				0.12c	
				0.53c	
				0.37c	
				0.46c	
				0.25c	
				(10.79)	
				0.53d	
				0.35d	
				0.16d	
				0.65d	
				(37.11)	

Note: the mean and standard deviation is shown (in parenthesis). BF_{01}^{x} – BFDA ratio when assumptions are violated are shown below F-values (also in parenthesis). [a] – F and BF_{01}^{x} values for stimuli when humor pre-disposition covariate is controlled. [b] – F and BF_{01}^{x} values for stimuli when prior-knowledge covariate is controlled. [c] – F and BF_{01}^{x} values for stimuli when digit-span covariate is controlled. [d] – F and BF_{01}^{x} values for stimuli when arrow-span covariate is controlled.
The variances were non-homogenous ($p < .05$), thereby robust Welch analysis in SPSS and R-WRS2 package were used for these DVs. Lastly, just like when the p value is conventionally written as $p < 0.01$ even if smaller, the BF_{01} values that were also smaller than 0.01 were just written as such. These measures are taken for the reader's convenience.
*$p = < .10$. ** $p = < .05$. *** $p = < .001$.

Thus in this study, if the majority of the trimmed means were found to be significantly different, the whole IV, covariate, and DV relation was also considered to be so (and vice versa). Lastly, the package could not calculate the independent contribution of each covariate on a DV. Nevertheless, the reader is urged to appreciate the insight achieved with these results, given the limitations of our current technology on analyzing with robust ANCOVA for DVs with non-homogenous variances. At least the technology is available with such limitations; it is an advantage, considering that there is currently no robust BFDA analysis; the reader is cautioned to interpret BF_{01}^{x} as revealed in results in Table 2.

4.2. Findings

Having said that, the results in Table 2 show that the manipulation was successful. The DVs humorous scale and mirth were highly significantly different and in favor of the IHCALM condition, when analyzed with robust ANOVA in SPSS. When each of the covariates was controlled separately, the robust ANCOVA in R-WRS2 was also significant for nearly all trimmed means.

Further, there was near zero evidence in favor of the null hypothesis, according to BFDA. By looking at the NHST and BFDA results in Table 2, we can use the cut-off point set in Table 3 and see that the data were nearly always in favor of "*Evidence for H_1*" outcome for humorous scale and mirth DVs (Tables 2 & 3). Thus, both the manipulations were successful and participants found the IHCALM condition significantly more funny than the NH condition, even when humor pre-disposition, prior knowledge, digit-span, or arrow-span were controlled.

Table 3. Four general outcomes between the null hypothesis testing (NHST) and the Bayesian factor design analysis (BFDA) test.

	NHST ($p < 0.05$)	**NHST ($p > 0.05$)**
BFDA (<3)	Evidence for H_1	No power
BFDA (>3)	N/A or Humorous	Evidence for H_0

Source: Dorambari (2022).

The remaining DVs with non-homogenous variances were found to be mostly non-significant. The robust Welch ANOVA in SPSS initially showed that the difference between the means of the cognitive load error DV was significant and in favor of the NH condition. Since the BFDA ratio was also low, initially the data favored the "*Evidence for H_1*" outcome. However, when the covariates were controlled, cognitive load error was found to be both not significant and the BFDA ratio was mostly less than 3, which indicated that the data was largely in favor of the "*No Power*" outcome (Tables 2 & 3).

On the other hand, the negative academic emotion afraid had non-significant NHST values as well as the BFDA ratio larger than 3. The data mostly favored the "*Evidence for H_0*" outcome, which was the same outcome when the covariates were controlled (Tables 2 & 3). Thereby, the results in this robust analysis showed that the participants in the IHCALM and NH conditions were neither afraid nor experienced any significantly different extrinsic cognitive load. Thus the analysis continued with DVs with homogenous variances.

Since the other cognitive load DVs had homogenous variances and regression slopes, the MANOVA was initially run in SPSS. Results showed that the multivariate statistic was not significant, V = 0.02, F (3, 224) =1.56, p = 0.20. However, there were significant differences found in the follow up ANOVA.

More precisely, the differences between the means were initially significant and in favor of the NH condition for the cognitive load misplaced digits DV. Further, the BF_{01} value was 1.06, which indicated that the data favored the "*Evidence for H_1*" outcome (see Table 3 and 4). However, when the covariates were controlled, the means were no longer significantly different and the BF_{01} value was only 0.49, which indicated that the data now favored the "*No power*" outcome (just like with the cognitive load error DV in Table 2).

The remaining DVs of cognitive load missing values, similar repeating values, repeats, negative academic emotions (NAE), positive academic emotions (PAE), BAS, correct answers, missing answers, metacognitive percent, and missing metacognitive values were not significant and mostly had BFDA ratios higher than 3. With such results, the data largely favored the "*Evidence for H_0*" outcome (Tables 3 & 4). This again demonstrates that the IHCALM condition is similar to the NH condition. For all other "*Evidence for H_1*" outcomes that came from the independent contribution of covariates, the reader is directed to Table 4.

One surprising exception to the above findings is the PAE interest, which was found to have significantly different means, which had a low BFDA ratio, and a medium effect size, all in favor of the NH condition, regardless of whether covariates were controlled or not (Tables 3 & 4). Although the IHCALM condition had humor while the NH condition did not, this finding was surprising because the participants found the latter more interesting, nevertheless. Also surprising was that the BF_{01} surpassed the cut-off point in Table 3, which indicates that the data favored the "NA" or the "Humorous" outcome.

Table 4. The mean, standard deviation, ANOVA and ANCOVA F-values, BFDA, and effect sizes for dependent variables that had homogenous variances and regression slopes.

Dependent Variables	NH M (SD)	IHCALM M (SD)	F_{ANOVA} (BF_{01})	F_{ANCOVA} (BF_{01})	Cohen's d
Cognitive Load Misplaced Digits	2.80	3.75	4.0*	3.54s	0.27
	(3.13)	(3.91)	(1.06)	(0.49s)	
				1.61a	
				(4.89a)	
				0.38b	
				(4.76b)	
				11.75c***	
				(0.02c)	
				0.12d	
				-1.53	
Cognitive Load Missing Values	2.53	2.00	0.52	0.84s	0.08
	(8.2)	(4.07)	(5.4)	(74.95s)	
				0.01a	
				(6.61a)	
				1.71b	
				(3.48b)	
				0.04c	
				(6.17c)	
				5.71d***	
				(0.06d)	
Cognitive Load Repeated Similar Values	0.43	0.68	0.67	0.41s	0.15
	(1.36)	(1.84)	(5.05)	(56.88s)	
				0.01a	
				(6.9a)	
				0.08b	
				(6.9b)	
				0.72c	
				(2.94c)	
				2.07d	
				(1.32d)	
Repeats	0.77	0.62	0.43	0.40s	
	(1.65)	(1.62)	(5.65)	(1040.10)	
				0.01a	
				(6.91a)	
				0.09b	
				(6.43b)	
				1.29c	
				(3.63c)	
				0.02d	
				(6.58d)	
Distress	1.94	1.79	0.2	0.15s	0.12

	(1.19)	(1.21)	(6.3)	(33.36s)	
				6.95a***	
				(0.52a)	
				0.12b	
				(6.29b)	
				3.98c*	
				(2.34c)	
				0.54d	
				(6.91d)	
Upset	1.76	1.95	1.87	2.14s	0.19
	(0.99)	(1.06)	(2.86)	(37.91s)	
				5.79a*	
				(0.66a)	
				0.33b	
				(6.06b)	
				2.10c	
				(4.67c)	
				0.81d	
				(6.80d)	
Scared	1.50	1.56	0.43	0.54s	0.07
	(1.00)	(0.85)	(5.65)	(642.99s)	
				0.22a	
				(6.71a)	
				0.62b	
				(5.24b)	
				0.75c	
				(6.30c)	
				1.91d	
				(4.27d)	
Ashamed	1.45	1.64	3.21	2.98s	0.21
	(0.87)	(0.93)	(1.53)	(87.30s)	
				0.17a	
				(5.95a)	
				0.01b	
				(6.67b)	
				2.86c	
				(1.10)	
				0.16d	
				(3.71)	
Hostile	1.51	1.44	0.21	0.29s	0.08
	(0.97)	(0.8)	(6.27)	(116.94s)	
				2.53a	
				(3.13a)	
				0.16b	

				(6.45b)	
				4.40c*	
				(0.99c)	
				0.01d	
				(5.42d)	
Nervous	1.96	2.13	1.49	1.70s	0.14
	(1.13)	(1.18)	(3.42)	(246.61s)	
				2.05a	
				(3.38a)	
				0.01b	
				(6.92b)	
				1.88c	
				(4.6c)	
				1.11d	
				(6.36d)	
Guilty	1.50	1.39	0.51	0.66s	0.14
	(0.88)	(0.8)	(5.44)	(1466.47s)	
				0.13a	
				(6.48a)	
				0.44b	
				(5.83b)	
				0.04c	
				(6.86c)	
				0.10d	
				(6.74d)	
Jittery	2.17	2.30	1.49	2.18s	0.12
	(0.92)	(1.14)	(3.43)	(7.79s)	
				0.87a	
				(3.41a)	
				2.78b	
				(2.65b)	
				7.29c***	
				(0.21c)	
				0.29d	
				(6.67d)	
Irritate	1.62	1.62	0.01	0.01s	0.00
	(0.95)	(1.01)	(6.90)	(39.96s)	
				5.09a*	
				(0.95a)	
				0.94b	
				(4.53b)	
				3.84c*	
				(1.56c)	
				0.01d	

				(5.09d)	
Strong	2.84	2.78	0.22	0.32s	0.05
	(1.15)	(1.11)	(6.23)	(1151.39s)	
				0.42a	
				(5.34a)	
				0.41b	
				(5.97b)	
				0.54c	
				(5.14c)	
				0.03d	
				(6.92)	
Interest	3.20	2.69	10.89***	10.12s***	0.41
	(1.25)	(1.25)	(0.04)	(1.20s)	
				0.03a	
				(6.89a)	
				0.01b	
				(6.61b)	
				2.31c	
				(1.63c)	
				0.33d	
				(2.45d)	
Proud	2.28	2.25	0.03	0.46s	0.03
	(1.14)	(1.12)	(6.81)	(3.94s)	
				2.21a	
				(1.32a)	
				0.13b	
				(6.26b)	
				2.31c	
				(5.40c)	
				10.82d***	
				(0.10d)	
Alert	3.07	2.95	0.37	0.52s	0.10
	(1.24)	(1.17)	(5.81)	(210.92s)	
				0.05a	
				(5.97a)	
				1.15b	
				(4.01b)	
				2.40c	
				(3.59c)	
				2.71d	
				(3.94d)	
Inspire	2.82	2.73	0.54	0.61s	0.07
	(1.15)	(1.2)	(5.36)	(395.68s)	
				2.99a	

				(1.53a)	
				0.23b	
				(5.92b)	
				0.01c	
				(6.87c)	
				0.17d	
				(6.40d)	
Determined	2.59	2.50	0.3	0.82s	0.07
	(1.11)	(1.16)	(6.04)	(45.69s)	
				1.17a	
				(5.29a)	
				0.23b	
				(5.84b)	
				1.86c	
				(6.41c)	
				7.53d**	
				(0.53d)	
Excited	3.20	3.07	0.47	0.51s	0.11
	(1.13)	(1.18)	(5.53)	(1865.79s)	
				0.03a	
				(6.86a)	
				0.08b	
				(6.76b)	
				0.02c	
				(6.83c)	
				0.01d	
				(6.91d)	
Enthusiastic	2.54	2.38	1.01	0.84s	0.15
	(1.15)	(1.08)	(4.29)	(130.48s)	
				1.25a	
				(2.49a)	
				0.27b	
				(6.23b)	
				3.29c	
				(1.41c)	
				0.59d	
				(6.46d)	
Attentive	3.52	3.50	0.06	0.16s	0.02
	(1.13)	(1.14)	(6.73)	(409.49s)	
				0.04a	
				(6.85a)	
				0.01b	
				(6.91b)	
				3.46c	

				(2.29c)	
				1.77d	
				(5.30d)	
Active	3.52	3.41	0.68	0.10s	0.12
	(0.93)	(1)	(5.02)	(463.84s)	
				0.10a	
				(6.90a)	
				0.34b	
				(6.06b)	
				2.49c	
				(3.20c)	
				1.10d	
				(6.11d)	
BAS	0.72	0.77	0.02	0.03a	0.02
	(2.48)	(2.52)	(6.73)	(919.50s)	
				0.27a	
				(3.89a)	
				0.48b	
				(5.39b)	
				0.55c	
				(6.0c)	
				0.86d	
				(6.27d)	
Correct Answers	13.17	12.25	1.71	0.97s	0.20
	(4.7)	(4.29)	(2.77)	(0.01s)	
				1.77a	
				(4.15a)	
				0.03b	
				(5.86b)	
				13.15c***	
				(0.01c)	
				10.84d***	
				(0.01d)	
Missing Answers	0.38	0.41	0.63	0.04s	0.03
	(1.59)	(1.28)	(6.74)	(86.20s)	
				0.28a	
				(5.69a)	
				2.38b	
				(1.85b)	
				0.03c	
				(6.26c)	
				4.16d*	
				(0.63d)	
Metacognitive Percent	51.31	53.57	1.71	3.38s	0.14

	(16.39)	(16.34)	(3.09)	(0.01s)	
				4.01a*	
				(1.98a)	
				1.38b	
				(3.50b)	
				8.91c**	
				(0.02c)	
				5.89d*	
				(0.03d)	
Metacognitive Missing Values	0.62	0.41	0.63	1.42s	0.11
	(2.36)	(1.45)	(5.13)	(73.04s)	
				0.01a	
				(6.65a)	
				2.67b	
				(1.99b)	
				2.07c	
				(4.86c)	
				3.5d	
				(2.38d)	
TP	8.26	8.33	0.19	0.58s	0.02
	(4.13)	(3.71)	(6.31)	(303.02s)	
				2.91a	
				(1.34a)	
				0.06b	
				(6.87b)	
				0.10c	
				(3.52c)	
				0.05d	
				(6.92d)	
TN	4.69	6.34	4.92*	3.86s*	0.36
	(3.75)	(5.25)	(0.7)	(1.07s)	
				6.12a*	
				(0.50a)	
				2.67b	
				(1.42b)	
				0.92c	
				(5.11c)	
				0.03	
				(4.27d)	
FP	13.95	12.73	2.08	1.37s	0.23
	(4.97)	(5.81)	(2.6)	(39.46s)	
				4.43a*	
				(0.89a)	
				2.35b	
				(1.97b)	
				0.18c	
				(6.49c)	
				0.20d	
				(4.28d)	
FN	2.31	2.23	0.56	1.25s	0.03
	(2.46)	(2.52)	(5.33)	(0.27s)	

				14.32a***	
				(0.01a)	
				1.03b	
				(5.08b)	
				2.32c	
				(1.97c)	
				1.40d	
				(4.61d)	

Note: The means are shown for each condition together with the standard deviation (in parenthesis). BF_{01} – Bayesian factor analysis results are shown below F-values (also in parenthesis). [s] – F and BF_{01} of stimuli values when covariates are controlled. [a] – F and BF_{01} values for humor pre-disposition covariate. [b] – F and BF_{01} [x] values for prior knowledge covariate. [c] – F and BF_{01} [x] values for digit-span covariate. [d] – F and BF_{01} [x] values for arrow-span covariate. The variances and regression slopes were homogenous ($p > .05$).
*p = < .10. ** p = < .05. *** p = < .001.

The author of this study calls this outcome as "Humorous" because it may initially imply that there is strong evidence both for H_1 and H_0 at the same time, which can be initially conflicting, unsettling, and can create inner tension. However, the tension is loosened upon realizing that the outcome occurred because of an error, small sample, small cut-off boundary, or because finding evidence for H_0 is more difficult than for H_1 in general (Stefan, Gronau, Schönbrodt, & Wagenmakers, 2019). In such a state, the tension release may also result in exhilaration for the researcher, hence the "Humorous" name for this outcome. For this study, it just meant that since the PAE interest was significantly different after the covariates were controlled, there was a moderate evidence in favor of the null hypothesis, the data point to the borderline "*Evidence for H_1*" outcome.

Since the metacognitive missing values in the IHCALM condition were highly similar to the NH condition (BF_{01} = 73.04), another pleasant exception in the current study occurred, that of additional analysis. The current study analyzed every metacognitive unit separately and compared their means for differences between the stimuli conditions (see Table 5 for metacognitive units).

Table 5. Four units of the variable metacognitive percent.

	Correct Answer	Wrong Answer
Certain	True Positive (TP)	False Positive (FP)
Uncertain	False Negative (FN)	True Negative (TN)

On this additional analysis, the metacognitive TN DV was found to have a significant different mean that was in favor of the IHCALM condition and had a medium effect size as well. With a low BF_{01} ratio, the data initially was in favor of the "*Evidence for H_1*" outcome. The outcome changed to the borderline "*Evidence for H_1*" outcome, when the covariates were controlled (Tables 3 & 4). This means that compared to the participants in the NH condition, the ones in the IHCALM condition were more certain that their chosen answers were wrong during tests, even when the covariates were controlled.

5. DISCUSSION

To summarize, the purpose of this study was to demonstrate that the IH as an IV in IHCALM did not harm learning. The method chosen to demonstrate such outcomes was achieved by direct replication and controlling the covariates. As such, it was found that the only DVs that favored the "*Evidence for H_1*" outcome was cognitive load error, cognitive load misplaced digits, humorous scale, mirth, interest, and TN; all other DVs had data that were mostly in favor of the "*Evidence for H_0*" outcome, while some data pointed to the "*No Power*" outcome. Thereby, it was found that IHCALM was a new CATLM approach, not a seductive detail, and can be harmlessly used in education.

However, even when the cognitive load DVs were initially significantly different, the outcome later had changed. After the covariates were controlled, the cognitive load DVs were no longer significantly different nor had a lower than 3 BFDA ratio. The resulting data now both favored the "*No Power*" outcome and hinted that the covariates (not the stimuli) were responsible for the means being as significantly different in the first place.

In the follow up ANCOVA, it was revealed that the covariate digit-span had a significant and small effect on the amount of cognitive load misplaced digits, which were in favor of the NH condition (see Table 4). Parameter estimates revealed that the *b*-value was negative at β = -0.09, $t(225) = -3.43$, p <.01, Partial - $\eta^2 = 0.05$. This means that for roughly every 10 points that the participants scored less in the digit-span test, they inserted 1 digit in its wrong place during preload, which mostly happened after watching an IHCALM presentation.

On the other hand, the participants' arrow-span (that measured the visuospatial sketchpad) also had a significant and a small effect on cognitive load missing values, however now this was in favor of the IHCALM condition. Parameter estimates again showed that the value was negative at β = -0.55, $t(225)$, = -2.39, p < .01, Partial - $\eta^2 = 0.03$. This means that for nearly every 2 errors that the participants made during the arrow-span test, they did not insert a value at all during preload, which mostly occurred when watching the NH presentation.

Thereby, the less the digit-span the more the participants that watched the IHCALM condition would experience extraneous cognitive load and placed the digits in the wrong places. On the other hand, the less was the arrow (visuospatial sketchpad) span, the more participants watching the NH condition would experience extraneous cognitive load and not insert a value at all. It seemed that cognitive load outcomes depended on the students WMC level and type, which has practical implications. Although the independent contribution of the WMC covariates was responsible for the DVs being as significantly different (not stimuli as previously mentioned), it may be worthy to re-assert here that the effect size of parameter estimates was low (Partial - $\eta^2 = 0.03 - 0.05$).

Other cognitive load findings related to intrinsic load were measured with the DV repeats. Although the IHCALM condition had both more words and video sequences to account for IH, there was no significant difference of intrinsic load. The reason why intrinsic load was similar between the two conditions might have to do with the personalization effect (Mayer, 2008), which is arguably found on every humor.

It seemed that students initially experienced more intrinsic load in the IHCALM condition than in the NH condition, due to the difference in video sequence amount. However, the load was compensated with less germane cognitive load due to the personalization effect (Mayer, 2008) of IH, which was only present in the IHCALM condition. As such, although initially the NH condition had less words and sequences (and probably less intrinsic load also), the final balance came up to be extremely similar for both conditions (BF_{01} = 1040.10).

However, the "*Evidence for H_1*" outcome for the PAE interest could not be easily explained. It is tempting to claim that the participants' limited capacities were not intrinsically loaded in the NH condition, since it had less words and sequences. With more WMC units to spare, the participants in the NH condition could appreciate more the same 3D imageries that were also used in the IHCALM condition to depict brain cells; thus felt more interest in the former rather than the later condition. However, this contrary to expectation outcome could not be explained by intrinsic load because it was measured and found to be extremely similar with the DV repeats (as mentioned above). Another tempting explanation could be extraneous cognitive load. Participants in the IHCALM condition initially experienced more extraneous cognitive load than the NH condition. This again may have spared the participants' WMC units and allowed them to use those to appreciate the same 3D imageries of neural activity, which caused them to experience more interest.

However, when the WMC covariates digit-span and arrow-span were controlled, there was no more statistical power to calculate the data and point to a conclusive direction, which is called the "*No power*" outcome (Tables 3 & 4). Future studies could explore why there may be more interest in the NH condition compared to the IHCALM condition by a replication study to see if it is a re-occurring pattern. If the pattern is re-occurring, then perhaps a qualitative study may better explore the reasons why this happens.

A similar borderline "*Evidence for H_1*" outcome was found with the metacognitive unit TN DV. The TN DV was significant in the initial MANOVA, remained significant when the covariates were controlled in the

MANCOVA, and had a medium effect size, as well as a BFDA ratio > 3, which pointed to the borderline "*Evidence for H_1*" outcome, just like with the PAE interest. One difference between these only two borderlines "*Evidence for H_1*" outcomes of interest and TN was that the former was in favor of the NH condition, while the latter favored the IHCALM condition. The second difference was that the latter can be better explained than the former.

The Wyer and Collins (1992) theory of humor has eight postulates, and those are briefly paraphrased here: 1) When exposed to a stimulus, MRs are created and stored in a particular location in memory during encoding; 2) During encoding, those MRs are created and stored in a particular way that one uses to understand, interpret, and later find easiest to retrieve from LTM; 3) Upon encoding, other associated MRs previously stored in one area in LTM may be retrieved and linked with the new coming MRs that are created during interpretation, which are then used to create general expectations and interpret future MRs that may be created related to the stimulus; 4) If the future stimuli related MRs do not fit with earlier expectations and interpretations in one area of the LTM, one attempts to create a different understanding and interpretation by using other MRs from other areas in their LTM, which if found leads to reinterpretation; 5) If the reinterpretation violates strongly established normative MRs in LTM, then a more consonant reinterpretation is sought to avoid dissonance; 6) Humor is elicited when those new MRs during reinterpretation that violated the established normative MRs are diminished/ridiculed/benignly violated from the person; 7) The amount of humor elicited depends on time and cognitive load that is required for reinterpretation; and 8) For the reward of exhilaration, the participants continue elaborating on the potential implications of the various MRs that they created and later diminished/ridiculed/violated during reinterpretation.

Thereby the participants in the IHCALM condition may have encoded more words, sequences, and IH in more areas on their LTM than the NH condition. This may have caused more associated MRs to be linked when interpreting the IH in the IHCALM presentation. Since the MRs in IHCALM benignly violated their previously held misconception MRs, this may have caused them to re-understand and reinterpret the information in the IHCALM presentation. This in turn allowed participants to diminish/ridicule/benignly violate their previous misconceptions; leading them to enjoy the exhilaration more than the participants in the NH condition.

Finally, for the sake of enjoying the humorous and rewarding exhilaration, they may have continued diminishing/ridiculing/benignly violating their earlier misconceptions. More precisely, they may have enjoyed diminishing misconceptions about neural cells by elaborating on the potential implications that their previous misconceptions could have on their lives. When this last postulate 8 was applied in IHCALM, it could mean that participants could also positively reinforce with exhilaration the desired new MRs from the IHCALM presentation (at the price of earlier misconceptions) into their LTM.

All these additional cognitive activities may have meta-cognitively benefited the participants in the IHCALM condition more than the NH condition in several ways, such as: a) They encoded the presentation in more areas in their LTM (i.e., one for instruction and the other area for humor, at least); b) They linked more associated MRs that came from more areas of their LTM during understanding and interpretation of IHCALM; c) They created more expectations and reinterpretations; d) They enjoyed exhilaration by diminishing/ridiculing/benignly violating their previous misconceptions; and e) They continued doing so for the pleasure of exhilaration, which may have also positively reinforced the information in IHCALM with their LTM.

As a result of this additional cognitive activity, the participants in the IHCALM condition although may have not been as certain as to know which answer is correct, they at least were more certain than the NH condition that their chosen answer is incorrect. They were more certain that their chosen answer was wrong because they may have been enjoying rewarding themselves with exhilaration. The exhilaration was achieved by diminishing/ridiculing/benignly violating the humorous implications of those chosen wrong answers/ misconceptions.

Further, according to the 8th postulate, the participants will continue rewarding themselves, even after the exams. This has other practical implications, since it is a desired learning outcome if information from IHCALM is stored and is more resilient to memory decay. A future study could test if information is retained longer in the IHCALM condition via repeated measure analysis.

6. IMPLICATIONS

The most important practical implication of this study is that the multimedia instructional designer (MID) could use IHCALM (rather than an NH presentation), if they wish their students to be more certain of what they do not know during exams. It may also be possible that the students will retain the learned knowledge for longer with IHCALM, since they may continue enjoying themselves with exhilaration by diminishing/ridiculing/benignly violating their previously held misconceptions. With more exhilaration, the students could reward themselves by linking the IHCALM information with their LTM.

Another practical implication for MID that wishes to use IHCALM is taking into consideration the class WMC type and level. If the digit-span is too low, the students may miss more information in the IHCALM presentation. If the visuospatial sketchpad is too low, the students may lose more information in the NH presentation. However, since the effect size of such individual WMC differences were found to be low, it should not be too much of a trouble if a MID wish to use IHCALM nevertheless. The theoretical implications of these findings are wider. The IHCALM approach may be the way forward in lowering the Dunning-Krugger effect (DKE) (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Basically, the DKE illustrates that students overestimate their knowledge and mostly claim that they have scored above average in the exams, when they have not (i.e., they fall in the FP metacognitive unit category, see Table 5). If it is the aim of calibrating the students' metacognitive certainty with their actual performance, IHCALM was found to influence students to be more certain of what they do not know (i.e., moved participants from the FP category to the TN one) to a significant degree. Arguably, calibrating students' metacognition with their performance and moving them away from the FP to the TN category is less desirable than moving them from the FP to the TP one. However, the role of the effects of humor in general (and IHCALM in particular) on calibration and lowering the DKE may be the future step that remains to be studied.

7. CONCLUSION

It has been demonstrated both with direct replication and covariate control in this study that IHCALM is not a seductive detail effect that harms learning. Even though less interesting, the future MID could now design an IHCALM presentation to use in education and benefit the students with humorous mirth and learning, which may also help them retain knowledge longer and do not succumb to the effects of memory decay. Although with small effect sizes, the future MID could also benefit from this study by basing their decision whether they should design an IHCALM or NH presentation, according to their student' MRs as well as their WMC types and levels.

On top of such benefits, IHCALM was also found to increase the students' metacognition and made them more certain of what they did not know during exams. This finding seems promising for metacognitive calibration and lowering of the DKE during exams. Future studies should look more into the positive effects of humor in general and IHCALM in particular in lowering the DKE of students both in education and possibly beyond. Lowering the DKE and increasing metacognitive calibration could help them make more accurate estimations of their current performance; based on such estimations, they could also make more accurate decisions in their future lives.

Funding: This study received no specific financial support.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

REFERENCES

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.Available at: https://doi.org/10.1146/annurev-psych-120710-100422.

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.Available at: https://doi.org/10.1027/1618-3169.49.2.109.

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.Available at: https://doi.org/10.1037/0022-3514.67.2.319.

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.Available at: https://doi.org/10.3758/BF03194326.

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.Available at: https://doi.org/10.1016/j.concog.2009.09.009.

Dorambari, D. (2022). Instructional humour and cognitive affective learning with multimedia (IHCALM). *Prizren Social Science Journal, 6*(1), 18–29.Available at: https://doi.org/10.32936/pssj.v6i1.282.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83-87.Available at: https://doi.org/10.1111/1467-8721.01235.

Falk, D. R., & Hill, C. E. (1992). Counselor interventions preceding client laughter in brief therapy. *Journal of Counseling Psychology, 39*(1), 39-45.Available at: https://doi.org/10.1037/0022-0167.39.1.39.

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19-38.Available at: https://doi.org/10.1016/j.brat.2017.05.013.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 443.Available at: https://doi.org/10.3389/fnhum.2014.00443.

Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Journal of Social Psychology, 38*(1), 7-16.Available at: https://doi.org/10.1024/0044-3514.38.1.7.

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 Behavior, 44*, 81-95.Available at: https://doi.org/10.1016/j.chb.2014.11.009.

Kruley, P., Sciama, 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.Available at: https://doi.org/10.3758/BF03200853.

Mair, P., & Wilcox, R. (2018). Robust statistical methods using WRS2. The WRS2 Package. Retrieved from: https://www.academia.edu/39358727/Robust_Statistical_Methods_Using_WRS2?auto=citations&from=cover_page.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422-430.Available at: https://doi.org/10.1016/j.concog.2011.09.021.

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.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*(8), 760-769.Available at: https://doi.org/10.1037/0003-066x.63.8.760.

Mayer., R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33*, 12-18.Available at: https://doi.org/10.1016/j.learninstruc.2014.02.004.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309-326.Available at: https://doi.org/10.1007/s10648-007-9047-2.

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44*, 267-278.Available at: https://doi.org/10.1016/j.chb.2014.10.061.

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), APA educational psychology handbook. Individual differences and cultural and contextual factors (Vol. 2, pp. 3–31): American Psychological Association.

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, *29*, 128-140.Available at: https://doi.org/10.1016/j.learninstruc.2013.02.006.

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry*, *18*(1), 1-8.Available at: https://doi.org/10.1186/s12888-018-1761-4.

Ruch, W. (1993). Exhilaration and humour. *Handbook of Emotions*, *1*, 605-616.Available at: https://doi.org/10.5167/uzh-77841.

Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction*, *44*, 65-73.Available at: https://doi.org/10.1016/j.learninstruc.2016.03.002.

Schneider, S., Nebel, S., Beege, M., & Rey, G. D. (2018). Anthropomorphism in decorative pictures: Benefit or harm for learning? *Journal of Educational Psychology*, *110*(2), 218-232.Available at: https://doi.org/10.1037/edu0000207.

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.Available at: https://doi.org/10.1007/s10648-011-9168-5.

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on bayes factor design analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042-1058.Available at: https://doi.org/10.3758/s13428-018-01189-8.

Stone, J. M., & Towse, J. (2015). A working memory test battery: Java-based collection of seven working memory tasks. *Journal of Open Research Software*, *3*(4), e5.Available at: https://doi.org/10.5334/jors.br.

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, *104*(2), 485–498.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Boutin, B. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58-76.Available at: https://doi.org/10.3758/s13423-017-1323-7.

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.Available at: https://doi.org/10.1037/0022-3514.54.6.1063.

Wyer, R. S., & Collins, J. E. (1992). A theory of humor elicitation. *Psychological Review*, *99*(4), 663.Available at: https://doi.org/10.1037/0033-295x.99.4.663.