ISSN 2630-516X (Online)
Volume 1 Issue 1 · December 2018

Journal of International Education and Practice

Bilingual Publishing Co.

Volume 1 Issue 1 • December 2018 • ISSN 2630-516X (Online)

Journal of International Education and Practice

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Volume 1 | Issue 1 | December 2018 | Page 1-55
Journal of International Education and Practice

Contents

Article

Copyright

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep

ARTICLE

The Use of Computer-Assisted Keyword Technique in Learning Pronunciation of English Weak Forms

Hassan Saleh Mahdi

Department of English, College of Arts and Sciences, University of Bisha, Balqarn, 61922, Saudi Arabia

ARTICLE INFO

Article history:
Received: 1 November 2018
Accepted: 5 December 2018
Published: 12 December 2018

Keywords:
Pronunciation
Keyword technique
Weak forms
CAPT
Dual coding theory

ABSTRACT

Keyword technique has been investigated extensively in the literature to examine its effectiveness on second language pronunciation. However, previous studies have not examined how computer-assisted keyword technique could aid second language learner to pronounce weak forms in English accurately. Therefore, this study aims at investigating the effect of computer-assisted keyword technique on acquiring weak forms pronunciation. Sixty Arab learners who learn English as a foreign language participated in the study and assigned into three groups. The first group viewed a video containing weak forms in English dubbed with the first language keyword. The second group was taught weak forms and aided with keyword traditionally. The third group was taught the same content using the traditional way without the keyword. Results of the post-tests indicated that participants in the keyword technique (either computer-assisted or traditionally) outscored the participants who did not use keyword technique. The study concluded that keyword technique either used with the help of a computer or used traditionally, is a useful tool to enhance weak forms pronunciation.

1. Introduction

Pronunciation is a key element in language learning. In addition to other aspects of language such as grammar and vocabulary, pronunciation plays a key role in determining the extent to which mutual understanding occurs among communicators. Regarding this, Lee, Jang, and Plonsky (2015) concluded that pronunciation instruction had a statistically significant impact on language learners[1]. Despite the importance of pronunciation, pronunciation training is mostly ignored in language classes[2]. It should be noted that the goal of teaching pronunciation is not to make learners sound like native speakers of the target language, but rather to enable learners to pronounce the target language in an intelligible way[3][4].

Technology offers several benefits for pronunciation instruction, such as the possibility to receive immediate feedback, individualized instruction, and access to an authentic input. Research need to explore the extent to which tools can be suitable for a given group of learners at a given point in time[5]. This study explores the potential of computer-assisted keyword technique to help English Foreign Language (EFL) learners improve their pronunciation of English weak forms.

**Corresponding Author:*
Hassan Saleh Mahdi
Department of English,College of Arts and Sciences, University of Bisha, Balqarn, 61922*, Saudi Arabia*
Email: almatari7@gmail.com

 DOI: https://doi.org/10.30564/jiep.v1i1.255

2. Literature Review

2.1 Pronunciation and Foreign Language Learning

Pronunciation is considered one of the primary features when assessing oral communication and spoken proficiency[6]. Indeed, pronunciation is important since mispronunciation may cause misunderstanding. Pennington and Richards (1986) defined pronunciation as the way to utter individual sounds, stress, and intonation patterns of the target language[7]. The purpose of the development of a foreign language learners' pronunciation is not only to imitate one norm of native speakers; but rather to achieve real communication and intelligibility between speakers[8]. The tendency nowadays is to achieve the highest degree of comprehensibility, since sounding like a native speaker is rare and often difficult[9] Comprehensibility, as defined by Derwing and Munro (2005), is a listener's perception of how difficult it is to understand an utterance and comprehensible pronunciation is considered as the chief goal of pronunciation training[10][11] since perfect pronunciation is sometimes impossible and not necessary.

Pronunciation remains a challenging task that many language learners face. Teaching pronunciation is often neglected compared to other skills such as reading and writing in many educational contexts. This is due to the fact many teachers lack enough training in this area, including how to describe intonation and stress of various lexical items[12].

2.2 Keyword Technique

Keyword technique is a technique that is designed to make a strong connection between the unknown word and its meaning. It is defined by Richards and Schmidt (2002) as "a learning strategy in which the learner thinks of a homophone (the keyword) in the native language for the word he or she is trying to remember in the target language"[13]. This technique is basically composed of two steps. First, the new word in L2 is linked to a word (a keyword) in the learner's first language which sounds like a part or the whole of the new word. Second, learner needs to think of a visual image where the meaning of the new word and the meaning of the keyword is combined[14]. Atkinson, Raugh and Schupbach (1975) proposed the following criteria for keywords: (a) the keyword sounds as much as possible like a part (not necessarily all) of the foreign word; (b) it is easy to form a memorable imagery link connecting the keyword and its English translation and (c) the keyword is unique (different from other keywords used in the vocabulary[15].

Keyword technique has a positive impact on both immediate and long-term retention. Beaton, Gruneberg, and Ellis (1995) concluded that even after ten years without opportunity for use, some memory for words learned by keyword technique remains[16]. Keyword technique is more powerful and easier to be used. Wang and Thomas (1995) found that the keyword technique took less time and gave better results for immediate recall and on a long-term measure[17]. Keyword technique as a way to enhance vocabulary learning has been examined by several studies[18][19][20][21][22][23][24][25]. The findings of these studies indicated that the keyword technique was effective, and learners' scores were higher in the case of experimental groups in both immediate post and delayed post-tests. The implementation of keyword technique has been examined for young children[26][27], teenagers[28], adults[29][23], and the elderly[31]. In addition, the effect of Keyword technique was examined to students with learning disabilities[31].

Although keyword technique has a lot of benefits, its effectiveness has been questioned in several studies. For example, Hall (1988) argued that the keyword technique can lead to poorer performances when it is used as the exclusive vocabulary learning method[32]. Other studies indicated that the keyword technique could produce similar or inferior results compared with traditional learning methods (rote learning), and that experienced language learners benefited less from the keyword technique than the inexperienced learners[33][34][35].

2.3 Computer-Assisted Pronunciation Training (CAPT)

Pronunciation learning can be enhanced with the help of a computer. The use of the computer in pronunciation learning can provide language learners with opportunities to practice L2 pronunciation in an effective and interesting way[36]. The use of computer has been recognized by many researchers as a beneficial tool for learning L2 pronunciation, as it has several characteristics that help L2 learners increase their performance in pronunciation. These characteristics can be categorized according to the following procedures: input, output, and feedback. Computer offers several opportunities that provide language learners with abundant input. Based on Input Hypothesis proposed by Krashen (1987), the basic component of successful language learning is input[37]. Learners need to be exposed to an abundant amount of L2 input. Using the computer to aid pronunciation learning can provide language learners with adequate authentic and contextualized spoken input.

Using CAPT enables language learners to access their own and others' pronunciation performance. It also offers a considerable promise for language pedagogy as a medium for improving language learners' productive and

DOI: https://doi.org/10.30564/jiep.v1i1.255

receptive competence in the pronunciation of the target language[38], as well as providing visual displays of various speech patterns[39]. In addition, CAPT has a range of advantages that create new opportunities for language learners. Firstly, it is quick in performing, analyzing and giving language learners feedback far faster than a teacher can do. Secondly, tasks can be repeated, and they are also precise and reliable in the sense of being the same every time. Pennington (1999) summarizes the properties of CAPT as the following: quick, repeatable, precise, reliable, authoritative, highly salient, multi-modal, individual and variable[38]. Luo (2016) found the CAPT particularly useful for reducing non-native English language learners' mispronunciation errors when compared to in-class only instruction on pronunciation[40]. Pi-Hua (2006) commented on the usefulness of CAPT for learners and teachers, with learners being able to practice independently and individually, while teachers can use it for assigning drilling practice, which is considered by some teachers to be overwhelming and time-consuming[41]. CAPT is also useful for providing language learners with correct and reliable feedback. Neri, Cucchiarini, and Strik (2008) tested the pedagogical effectiveness of automatic corrective feedback on segmental quality using CAPT[42]. The findings revealed an improvement in learners' pronunciation accuracy because of the training and spontaneous feedback they received. CAPT is also useful for enhancing speaking skills, as Fouz-González (2015) suggested that technology can assist learners in their productive practice[43]. One of the most well-known techniques is automatic speech recognition technology.

2.4 Weak Forms

Weak forms are one of two possible pronunciations for a word, in the context of the connected speech, the other being strong. The weak form is that which is the result of a word being unstressed, as in the normal pronunciation of the preposition of in cup of tea, and in most other grammatical words. Several words in English have more than one weak form such as (and /ænd/ can be /ənd/,/ən/, or /n/). In English, certain words can be pronounced in two different ways; strong form and weak form[44]. For example, students may not recognize /kəm n si:/ as come and see. This aspect is considered one of the difficulties that many Arab EFL learners face. It affects comprehending sentences containing weak forms[45]. Thus, it is important for EFL learners to be aware of weak forms.

2.5 Dual Coding Theory

Several theoretical rationales have been advanced to encourage the use of computer-assisted language learning. In his Dual Coding Theory (DCT), Paivio (1971, 1986) stated that the keyword method explicitly brings into play both verbal and imaginal processes[46][47]. He argued that the acoustical similarities between the first language (L1) keyword and the target word affects the development of vocabulary in the second language (L2) and direct relationships between L1 and L2 referents. He also argued that the use of mental images creates appropriate referential interconnections between L2 verbal representations and the imagery system.

Cognition according to DCT involves the activity of two distinct subsystems, a verbal system specialized for dealing directly with language and a nonverbal (imagery) system specialized for dealing with nonlinguistic objects and events. The systems are assumed to be composed of internal representational units, called logogens and imagens, that are activated when one recognizes, manipulates, or just thinks about words or things. The representations are connected to sensory input and response output systems as well as to each other so that they can function independently or cooperatively to mediate nonverbal and verbal behavior. The representational activity may or may not be experienced consciously as imagery and inner speech. The theory means that both systems are generally involved even in language phenomena.

2.6 Computer-assisted Keyword Techniques and Pronunciation

The combination of keyword technique with other strategies proved as efficient in aiding the information retention and recall over a period of time[48]. Therefore, using computer as an environment for keyword technique can be useful and create opportunities for language learners improve their achievement in language learning. The effect of the combination of keyword technique and computer in promoting EFL pronunciation has not yet thoroughly examined. The present study is an attempt to fill in this gap and explore the effect of using computer as an environment to present keyword for EFL pronunciation. The study seeks to answer the following research question: What is the effect of keyword technique on FL learners' pronunciation of English weak forms?

3. Methods

3.1 Research Design

This study follows the experimental design of the study in which three intact EFL university classes were randomly assigned to one of three groups: computer keyword technique, traditional keyword, and no keyword. The experiment was conducted for a month, as shown in Figure 1.

 DOI: https://doi.org/10.30564/jiep.v1i1.255

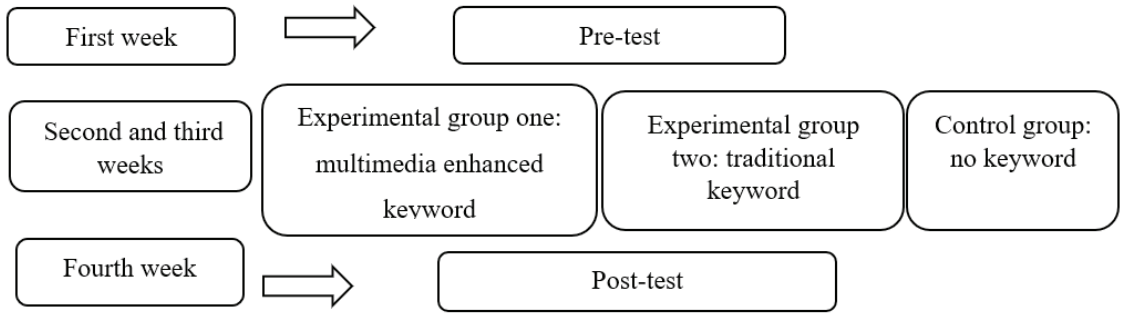

Figure 1. Study Design

In this study, the Keyword is the dependent variable and the independent variable is the use of computer as an aided technique to the Keyword. The dependent variable is a variable, which is observed to determine what effect the other types of variables may have on it. The dependent variables are the conditions or characteristics that appear, disappear, or change as the experimenter introduces, removes, or changes the independent variable.

3.2 Participants

The participants in the present study were 60 students from three intact EFL classes from three different colleges of University of Bisha, Saudi Arabia. The study was conducted in the first semester in the academic year 2018/2019. Their ages ranged from 19 to 24 years old. Twenty students were fe-male, and 40 students were male. Their average proficien-cy in English was categorized as intermediate based on the documents of the department. These students had been studying EFL for more than 7 years. Three experienced EFL instructors participated in the study. Each instructor used one different method to present the weak forms in his class.

3.3 Procedures

One class was held with the EFL instructors prior to the beginning of the study. In the first session, the pretest was distributed to all groups. The pre-test was administered in one regular session prior to the treatment. The participants were asked to read aloud ten sentences and record them. Their recordings were saved on the desktop in the language lab to be used later in the analysis. In the second and the third sessions, the weak forms were introduced (fourteen weak forms in each session). The same content was taught for all groups. However, participants in each group were trained with a different technique. Participants in the computer keyword group were instructed to watch a video which contained a speech of a native speaker of English. At any point where a word pronounced in a weak form, a short video clip with a keyword was added. The clips were added successively (i.e. the weak form in English was articulated and immediately a clip with a keyword was added). Participants in the traditional keyword group were taught weak forms using keywords traditionally. The instructor explained the weak form and presented a word that can be used as a keyword to enhance learning weak forms. No video was used to show the pronunciation of the weak forms in this group. Participants in the control group were taught weak forms traditionally. The instructor explained the weak form and pronounced them orally and the learners should repeat the correct way of pronunciation of weak forms. No video was used to show the pronunciation of the weak forms in this group. In the fourth session, the post-test was administered. The participants were asked to read aloud the same sentences as they appear in the pre-test and record them. Their recordings were saved on the desktop in the language lab to be used later in the analysis.

3.4 Testing Procedure

The test was developed by the researcher based on the sentences given in the textbook[44]. The pre-test consisted of ten sentences (see Appendix A). These sentences contained 22 words that can be pronounced either strong or weak form. These sentences were presented in a random order to each participant and making sure that participants did not realize the real purpose of the test which was intended to test their pronunciation of target words. To evaluate the participants' improvements in weak forms pronunciation, the participants were asked to read and record a set of ten isolated sentences. After two sessions, a post-test for the three groups was administered to test participants' pronunciation of the weak forms. Again, their responses were recorded for further analysis.

DOI: https://doi.org/10.30564/jiep.v1i1.255

3.5 Recording and Scoring

In the pre- and post-tests, the participants were asked to read the sentences aloud and record them. The recordings were made in the computer lab using head-mounted microphones. Head-mounted microphones were used to maximize the robustness of the recordings. The recording steps were performed as follows: First, the participants were asked to check the equipment by recording two short sentences which were not included in the tests. Second, the participants were asked to read aloud the sentences of the test. If the learner felt that he had not pronounced the sentences correctly, he or she could repeat them and record them as many times as needed. Third, each participant should record all the ten sentences and saved them in one file. By the end of the study, each participant would have two files: one for pre-test, and the second for the post-test. After recording, these files were given to three raters. The recordings were scored independently by three non-native speakers of English. Each rater was asked to provide a score of pronunciation quality of each sentence on a 5-point scale based on Isaacs, Trofimovich, and Foote (2017)[49].

Table 1. Isaacs et al. (2017) Comprehensibility global and analytical scale

Comprehensibility level	Overall description of comprehensibility
5	Speech is effortless to understand. Errors are rare and do not interfere with the message.
4	Speech requires little effort to understand. Errors minimally interfere with the message.
3	Speech requires some effort to understand. Errors somewhat interfere with the message.
2	Speech is effortful to understand. Errors are detrimental to the message.
1	Speech is painstakingly effortful to understand. Errors are detrimental to the message.
UR	Unable to Rate the speech.

The raters' scores were first analyzed to determine inter-rater reliability. A Cronbach's alpha coefficient of .88 was obtained, which can be considered good.

3.6 Data Analysis

Descriptive and inferential statistical procedures were performed in this study to determine the effect of keyword technique on EFL pronunciation. An ANOVA was conducted to determine the significance of the differences among groups and to which this significance referred to. An ANOVA describes a group of inferential statistical procedures which is used to analyze data from designs that involve two or more groups. Analysis of variance is a parametric statistical procedure for comparing two or more group means to see if there are any statistically significant differences among them.

The level of significance was .05 for all statistical analyses. These results will be discussed in the next section.

4. Results

To answer the research question of whether the students' pronunciation was improved as a result of the use of computer-assisted keyword technique, the mean and standard deviation across pre- and post-tests showed variation in the students' performance (Table 2). The use of Keyword technique was assigned as the between-subject variable, whereas the time of the tests was assigned as the within-subject independent variable. Levene test was used to examine the normality of the distribution. The results of the tests indicate that the data is normally distributed as p=.99 p >.05).

Results showed that a significant main effect of keyword technique existed from the pre-, and post-test designs. Results from ANOVA revealed that a significant main effect existed among the three types of keyword technique use. The computer-assisted keyword technique (CAK) and traditional keyword technique (TK) groups significantly outscored the no keyword technique (NK) group in the post-test (M = 39.3, 38.2, SD =3.37, 3.13 respectively). The performance of the CAK group over time was slightly higher than that of the TK group but the difference was not significant (MD =1.10, p=.273). The effect size was 1.9, which indicated a large effect of using the computer-assisted keyword in learning pronunciation.

Table 2. Descriptive Statistics of the Students' Performance of L2 Pronunciation Tests

	CAK (N = 20)		TK (N = 20)		NK (N = 20)		All Groups	
	M	SD	M	SD	M	SD	M	SD
Pre-test	12.8	2.78	12.6	2.72	12.48	2.56	12.48	2.66
post test	39.3	3.37	38.20	3.13	33.05	2.89	36.85	4.13

 DOI: https://doi.org/10.30564/jiep.v1i1.255

Table 3. ANOVA Results for L2 Pronunciation Tests

	Df	MS	F	P
Between groups	2	222.65	22.56	.000*
Within groups	57	9.86		.000*

The level of sig. = .05

Table 4. Tukey HSD Post-hoc Test

Dependent variables	(I) Keywords	(J) Keywords	Mean Differences (I-J)	Sig.
Keyword post-test	Computer KW	Traditional KW	1.10	.513
		No KW	625	.000
	Traditional KW	Computer KW	-1.10	.513
		No KW	5.15	.000
	No KW	Computer KW	-6.25	.000
		Traditional KW	-5.15	.000

The mean difference is significant at the of .05 level

Post-hoc pairwise comparisons using Tukey HSD yielded significant differences among the three groups over time (Table 4).

5. Discussion

This study sought to examine the effects of computer-assisted keyword technique on the pronunciation of English weak forms. Three groups participated in this study to find out the effect of computer-assisted keyword technique on the pronunciation of English weak forms. The results revealed that the positive impact of the keyword in both conditions (i.e. computer-assisted keyword or traditional use of keyword) on the pronunciation of weak forms. Both groups who used keywords as a technique to aid weak forms pronunciation outperformed the students in the group where no keyword was used. However, no statistically significant difference between computer-assisted keyword technique and traditional keyword technique, the scores of the students in the group used computer-assisted keyword technique are slightly higher. This indicates that computer-assisted keyword technique can be more useful in learning L2 pronunciation.

The study results are in line with Paivio's dual coding theory[46][47], which assumes that information is coded dually in the human mind either verbally (i.e., text and sounds) or non-verbally (i.e., picture and objects). These two systems are interconnected when words are represented by one system and can be activated by the other system or vice-versa (e.g., verbal activated by non-verbal). In this study, students' scores in computer-assisted keyword were slightly higher than the scores in traditional keyword because visual information (L1 keyword clip) was coded with verbal information (English pronunciation of weak forms). However, the students learning with no keyword mode performed worse on the pronunciation post-tests, as information was only presented traditionally (verbally). This study's findings demonstrate the significant benefits of using the keyword technique to assist English weak form pronunciation.

6. Conclusion

This study explored the effectiveness of learning pronunciation with keyword technique to enhance pronunciation improvement. In general, keyword technique had been found an effective tool to aid L2 pronunciation and it is more effective than no keyword used. However, the computer-assisted keyword was found slightly superior to the traditional keyword to facilitate weak forms pronunciation. The participants learned pronunciation better when they used computer-assisted keyword than when they used traditional keyword technique.

6.1 Pedagogical Implications

This study has generated several pedagogical implications for foreign language teachers.

L2 instructors might benefit from the results of this study to present material in a way that enhances pronunciation. One of the pedagogical implications for L2 instructors is to use both types of keywords (computer-assisted or traditional) to teach pronunciation. The teachers should make the keyword salient in the video presented to help the learners match the pronunciation of the weak forms accurately.

6.2 Limitations and Suggestions for Future Studies

There were several limitations to the current study. First,

DOI: https://doi.org/10.30564/jiep.v1i1.255

the small number of participants may threaten the validity of the study findings. Thus, a study with a large sample size would yield stronger evidence to generalize the findings. In this study, only one mode for a computer-assisted keyword was used (video L1 dubbing). Therefore, the study recommends that further studies investigate the impact of different modes (e.g. video + text that contains phonetic symbols) on learning pronunciation. The study suggests that further research explore the effect of computer-assisted keywords on learning pronunciation among young learners because age is a crucial factor for learning pronunciation.

References

[1] Lee, J., Jang, J. & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. Applied Linguistics, 36, 345–366. doi:10.1093/applin/amu040

[2] Nair, R., Krishnasamy, R. & De Mello, G. (2017). Rethinking the teaching of pronunciation in the ESL classroom. The English Teacher, 35, 27–40.

[3] Derwing, T. M., Munro, M. J., & Wiebe, G. (1997). Pronunciation instruction for fossilized learners. Can it help? Applied Language Learning, 8(2), 217–35.

[4] Derwing, T. M., Munro, M. J. & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. Language Learning, 48(3), 393–410. https://doi.org/10.1111/0023-8333.00047

[5] Jamieson, J., Chapelle, C., & Preiss, S. (2005). CALL Evaluation by Developers, a Teacher, and Students. CALICO Journal, 23, 93-138. http://dx.doi.org/10.1558/cj.v23i1.93-138

[6] Iwashita, N., Brown, A., McNamara, T. & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? Applied Linguistics, 29(1), 24–49. https://doi.org/10.1093/applin/amm017

[7] Pennington, M. C. & Richards, J. C. (1986). Pronunciation revisited. TESOL Quarterly, 20 (2), 207–225.

[8] Jenkins, J. (2000). The phonology of English as an International Language. Oxford, England: Oxford University Press.

[9] Kang, O. (2010). ESL learners' attitudes toward pronunciation instruction and varieties of English. Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference. Ames, IA: Iowa State University (pp: 105–118).

[10] Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research ‐ based approach. TESOL Quarterly, 39(3), 379-397. https://doi.org/10.2307/3588486

[11] Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. TESOL Quarterly, 25(3), 481–520.

[12] Couper, G. (2017). Teacher cognition of pronunciation teaching: Teachers' concerns and issues. TESOL Quarterly, 51(4), 820–843. doi: 10.1002/tesq.354

[13] Richards, J.C. and Schmidt, R. (2002) Longman Dictionary of Language Teaching & Applied Linguistics (3rd edition). Essex: Pearson Education Limited.

[14] Nation, I. S. P. (2001). Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

[15] Atkinson, R. C. & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. Journal of experimental psychology: Human learning and memory, 1(2), 126-133

[16] Beaton, A., Gruneberg, M., and Ellis, N. (1995). Retention of foreign vocabulary using the keyword method: a ten-year follow-up. Second Language Research, 11, 112-120. https://doi.org/10.1177%2F026765839501100203

[17] Wang, A. Y. & Thomas, M. H. (1995). Effect of keywords on long-term retention: help or hindrance? Journal of Educational Psychology, 87, 468-475.

[18] Atay, D., & Ozbulgan, C. (2007). Memory strategy instruction, contextual learning and ESP vocabulary recall. English for specific purposes, 26(1), 39-51. https://doi.org/10.1016/j.esp.2006.01.002

[19] Mahdi, H. S. & Gubeily, M. A. (2018). The Effect of Using Bizarre Images as Mnemonics to Enhance Vocabulary Learning. Journal of Social Studies, 24(1), 113-135. https://doi.org/10.20428/JSS.24.1.5

[20] McDaniel, M. A. & Pressley, M. (1989). Keyword and context instruction of new vocabulary meanings: effects on text comprehension and memory. Journal of Educational Psychology, 81, 204–213

[21] Pressley, M., Levin, J. R. & Miller, S. (1981). Cognitive strategy research: educational application. New York: Springer- Verlag

[22] Shapiro, A. M. & Waters, D. L. (2005). An investigation of the cognitive processes underlying the keyword method of foreign vocabulary learning. Language teaching research, 9 (2), 129-146. 10.1191/1362168805lr151oa

[23] Sagarra, N. & Alba, M. (2006). The key is in the keyword: L2 learning vocabulary learning methods with beginning learners of Spanish. The Modern Language Journal, 90, 228–243. https://doi.org/10.1111/j.1540-4781.2006.00394.x

[24] Taheri, A. A., & Davoudi, M. (2016). The effect of the keyword method on vocabulary learning and long-term retention. International Journal of Language and Linguistics, 3(1), 114-125. Available on http://ijllnet.com/journals/Vol_3_No_1_March_2016/10.pdf

[25] Tavakoli, M. & Gerami, E. (2013). The effect of keyword and pictorial methods on EFL learners' vocabulary learning and retention. Porta Linguarum: revista internacional

de didáctica de las lenguas extranjeras, (19), 299-316. Available on https://www.ugr.es/~portalin/articulos/PL_numero19/18%20%20Mansoor.pdf

[26] Avila, E. & Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. Language learning, 46(3), 379-395. https://doi.org/10.1111/j.1467-1770.1996.tb01241.x

[27] Wyra, M., Lawson, M. J. & Hungi, N. (2007). The mnemonic keyword method: The effects of bidirectional retrieval training and of ability to image on foreign language vocabulary recall. Learning and Instruction, 17(3), 360-371. https://doi.org/10.1016/j.learninstruc.2007.02.008

[28] Rodriguez, M. & Sadowki, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. Language learning, 50(2), 385-412. https://doi.org/10.1111/0023-8333.00121

[29] Beaton, A., Gruneberg, M., Hyde, C., Shufflebottom, A., & Sykes, R. (2005). Facilitation of receptive and productive foreign vocabulary learning using the keyword method: The role of image quality. Memory, 13(5), 458-471. https://doi.org/10.1080/09658210444000395

[30] Gruneberg, M. M. & Pascoe, K. (1996). The effectiveness of the keyword method for receptive and productive foreign vocabulary learning in the elderly. Contemporary Educational Psychology, 21(1), 102-109. https://doi.org/10.1006/ceps.1996.0009

[31] Mastropieri, M. A., Scruggs, T. E. & Fulk, B. J. (1990). Teaching abstract vocabulary with the keyword method: Effects on recall and comprehension. Journal of Learning Disabilities, 23(2), 92-96. https://doi.org/10.1177/002221949002300203

[32] Hall, J. W. (1988). On the utility of the keyword mnemonic for vocabulary learning. Journal of Educational Psychology, 80(4), 554

[33] Hogben, D. & Lawson, M. J. (1994). Keyword and multiple elaboration strategies for vocabulary acquisition in foreign language learning. Contemporary Educational Psychology, 19(3), 367-376.

[34] Van Hell, J. G. & Mahn, A. C. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. Language Learning, 47(3), 507-546.

[35] Willerman, B. & Melvin, B. (1979). Reservations about the keyword mnemonic. Canadian Modern Language Review, 35, 443-452.

[36] Mahdi, H. M. & Al-Khateeb, A. A. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis, Review of Education Journal.

[37] Krashen, S. D. (1987). Principles and practice in second language acquisition. New York

[38] Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. Computer Assisted Language Learning, 12(5), 427–440] https://doi.org/10.1076/call.12.5.427.5693

[39] Scarcella, R. C. & Oxford, R. L. (1994). Second language pronunciation: State of the art in instruction. System, 22(2), 221–230. https://doi.org/10.1016/0346-251X(94)90058-2

[40] Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. Computer Assisted Language Learning, 29(3), 451–476. https://doi.org/10.1080/09588221.2014.963123

[41] Pi-Hua, T. (2006). Bridging pedagogy and technology: User evaluation of pronunciation-oriented CALL software. Australasian Journal of Educational Technology, 22(3), 375–397. https://doi.org/10.14742/ajet.1292

[42] Neri, A, Cucchiarini, C. & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. ReCALL, 20(2), 225-243. https://doi.org/10.1017/S0958344008000724

[43] Fouz-González, J. (2015) Trends and Directions in Computer-Assisted Pronunciation Training. In: Mompean J.A., Fouz-González J. (Eds), Investigating English Pronunciation, pp. 314–342, Palgrave Macmillan, London, https://doi.org/10.1057/9781137509437_14

[44] Roach, P. (2009). English phonetics and phonology: A practical course, Cambridge University Press, United Kingdom.

[45] Smith, B. (2001). Arabic speakers. In: Swan, M., Smith, B. (Eds.), Learner English. Cambridge University Press, Cambridge (pp. 195—213)

[46] Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rinehart, & Winston. Reprinted 1979, Hillsdale, NJ: Lawrence Erlbaum Associate.

[47] Paivio, A. (1986). Mental representations: A dual coding approach. Oxford, UK: Oxford University Press

[48] Takac, V. P. (2008). Vocabulary Learning Strategies and Foreign Language Acquisition, Multilingual Matters Ltd

[49] Isaacs, T., Trofimovich, P. & Foote, J. A. (2017). Developing a user-oriented second language comprehensibility scale for English-medium universities. Language Testing DOI: https://doi.org/10.1177/0265532217703433

Appendix A (Sentences for the Pre- and Post-Test)

1- We can wait for the bus.
2- There are some books I must read.
3- She took her aunt for a drive.
4- The basket was full of things to eat.
5- You ought to have your own car.
6- He wants to come and see us at home.
7- Have you taken them from that box?
8- It's true that he was late, but his car was broken.
9- I shall take as much as I want.
10- I am too late to see him today.

DOI: https://doi.org/10.30564/jiep.v1i1.255

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep/

ARTICLE

Guessing and Nature of Multidimensionality Matter: A Cautionary Note on the Use of Fit Indices to Assess Unidimensionality of Binary Data

Yong Luo[1]* Dimiter M. Dimitrov[1,2]

1. National Center for Assessment in Higher Education, Riyadh, 12395, Saudi Arabia
2. George Mason University, Fairfax, Virginia, 22030, USA

ARTICLE INFO

Article history:
Received: 26 November 2018
Accepted: 13 December 2018
Published: 31 December 2018

Keywords:
Cut-off value
Model fit index
Guessing
Compensatory model
Noncompensatory model
Unidimensionality

ABSTRACT

Use of cutoff values for model fit indices to assess dimensionality of binary data representing scores on multiple-choice items is a popular approach among researchers and practitioners, and the commonly used cutoff values are based on simulation studies that used as the generating model factor analysis models, which are compensatory models without modeling guessing. Consequently, it remains unknown how those cutoff values for model fit indices would perform when (a) guessing exists in data, and (b) data follow a noncompensatory multidimensional structure. In this paper, we conducted a comprehensive simulation study to investigate how guessing affected the statistical power of commonly used cutoff values for RMSEA, CFA, and TLI (RMSEA > 0.05; CFA < 0.95; TLI < 0.95) to detect violation of unidimensionality of binary data with both compensatory and noncompensatory models. The results indicated that when data were generated with compensatory models, increase of guessing values resulted in the systematic decrease of the power of RMSEA, CFA, and TLI to detect multidimensionality and in some conditions, a small increase of guessing value can result in dramatic decrease of their statistical power. It was also found that when data were generated with noncompensatory models, use of cutoff values of RMSEA, CFA, and TLI for unidimensionality assessment had unacceptably low statistical power, and while change of guessing magnitude could considerably change their statistical power, such changes were not systematic as in the compensatory models.

1. Introduction

As one of the pivotal assumptions of item response theory (IRT), unidimensionality stipulates that item responses are driven by a single underlying latent variable. Numerous studies have shown that violations of the assumption of unidimensionality can lead to serious psychometric consequences such as biased item parameter estimates, equating errors, and misclassification

**Corresponding Author:*
Yong Luo,
National Center for Assessment in Higher Education, Riyadh, 12395*, Saudi Arabia*
Email: jackyluoyong@gmail.com

 DOI: https://doi.org/10.30564/jiep.v1i1.356

of examinees[1][2][3][4]. Aside from its critical importance in the valid application of IRT, the other two reasons that unidimensionality is of particular interest to researchers and practitioners, as summarized by Stout, are that the primary ability a test intends to measure should not be contaminated by other abilities and a unified latent variable is the precondition of meaningful comparison of individuals[5].

Probably due to its statistical importance and conceptual attractiveness, unidimensionality has received extensive attention in the psychometric literature[6][7][8][9] and a plethora of methods have been developed to assess unidimensionality[10][11][12][13][14][15][16][17][18]. Among them, factor analytic methods, "an important tool"[19] for dimensionality assessment of IRT models, are especially attractive to many structural equation modeling (SEM)[20] researchers since such methods are housed in the familiar SEM framework and require no additional IRT-based software programs other than common SEM ones. In this paper we focus on one such factor analytic method, namely the use of cutoff values of fit indices within SEM framework to assess unidimensionality with binary data. We assume that the binary data are item scores of multiple choice items and consequently, guessing is expected to exist within the data.

The use of fit indices for unidimensionality assessment makes intuitive sense in light of the mathematical equivalence between factor models with categorical variables and IRT[21][22][23][24][25]: if fit indices can be used to assess whether a one-factor model fits data satisfactorily, why cannot they be used to test whether a two-parameter normal ogive model, the IRT analog of the one-factor model with categorical variables, represents the data well? If these fit indices indicate good model fit based on some well-established cutoff values, it is concluded that the unidimensionality assumption is not violated. Despite its logical intuitiveness, this fit-index-based approach with binary data makes two implicit assumptions: first, those well-established cutoff values of model fit indices are applicable to cases of unidimensionality assessment with binary data; second, such cutoff values are robust to the existence of guessing.

The cutoff criteria for model fit indices proposed by Hu and Bentler[26] have been hugely popular among researchers interested in assessing the latent structure of their data. Although unidimensionality assessment is not included as a condition in their simulation study, these cutoff values have been nevertheless used by SEM researchers for unidimensionality assessment[27]. Despite their tremendous popularity, researchers have raised concerns about indiscriminate use of those indices[28][29][30][31][32], on the grounds that these fit indices are sensitive to different type of model misspecifications and consequently, establishment of cutoff values for model fit indices that are universally applicable is, if not impossible, very difficult. As pointed out by Huggins-Manley and Han[33], Hu and Bentler's simulation study[26], as well as other similar simulation studies[34][35] that address establishment of cutoff values, focus on misspecifications of factor loadings and/or latent variable correlations in multidimensional models. It remains unclear how model fit indices would perform when a unidimensional model is imposed upon data generated with multidimensional models.

While the impact of model misspecification type upon performances of model fit indices has been extensively studied in the literature, measurement quality, which can tremendously change the statistical behavior of model fit indices, fails to receive attention from researchers and practitioners with a few exceptions[36][37]. As demonstrated by McNeish, An, and Hancock, change of measurement quality (operationalized through the change of magnitude of standardized factor loadings) can result in drastically different distributions of model fit indices[36] and consequently, the model fit indices are meaningless without taking into consideration the standardized factor loadings. If measurement quality is conceptualized as the strength of the relation between indicators and the target latent variable, we argue that measurement quality can deteriorate with either the decrease of magnitude of standardized factor loadings, or with the introduction of guessing and the increase of guessing magnitude. As measurement quality can also be affected by the existence of guessing, we believe that the expected ubiquitous existence of guessing due to the common use of multiple-choice questions in educational setting, is of huge relevance when it comes to the dimensionality assessment of binary data. Since the common cutoff values of model fit indices were proposed in the factor analysis framework and based on data generating models that do not incorporate guessing, we believe their performances will be negatively affected by guessing based on previous studies that investigated the effect of guessing in factor analysis models[38][39][40]. To date, there have been no simulation studies that systematically investigate the guessing effect upon the performances of cutoff values of model fit indices.

Aside from guessing, another factor that has not received sufficient attention in the literature regarding the use of model fit indices for model assessment is the multidimensional nature of data (whether data exhibits compensatory or noncompensatory multidimensionality).

DOI: https://doi.org/10.30564/jiep.v1i1.356

While educational and psychological researchers usually focus on compensatory multidimensionality, noncompensatory multidimensionality occurs occasionally in some educational test items that require multiple skills and inadequacy in one skill cannot be compensated by other skills[41]. In the aforementioned simulation studies that dealt with categorical indicators[34][35], the researchers generated data based on factor analysis models that are equivalent to the compensatory IRT models, and it remains unknown how model fit indices, which are based on the factor analysis framework and therefore compensatory IRT models, will perform with data generated with noncompensatory IRT models. Previous studies have indicated that common dimensionality assessment methods that perform well with compensatory models may fall short with noncompensatory ones[42][43]. In addition, it is equally unclear whether and how guessing systematically affect the performances of cutoff values of model fit indices with data following noncompensatory structures.

The purpose of this study is to systematically investigate the impact of guessing upon the statistical power of cutoff criteria of model fit indices to refute unidimensionality when data are generated with both compensatory and noncompensatory IRT models. Specifically, since the standardized root mean square residual (SRMR) is not recommended for dichotomous items[35], we focus on the comparative fit index (CFI), the Tucker-Lewis Index (TLI), and the root mean square error of approximation (RMSEA), which are the also the fit indices reported in the popular latent variable modeling software Mplus[44] with the default weighted least squares mean- and variance-adjusted[45] estimator for categorical variables. Due to the existence of a large body of literature that provides excellent review of model fit indices, we do not review CFI, TLI, and RMSEA in this paper but refer interested readers to [46] for a comprehensive introduction[46].

The remainder of this paper is organized as follows. First, we review some influential simulation studies in which the commonly used cutoff values of fit indices were either established or validated. Next, we review previous studies that have investigated the effect of guessing in the factor analysis framework, followed by a review of studies dealing with dimensionality assessment of data generated with noncompensatory IRT models. In the method section we present two simulation studies conducted to investigate how commonly used cutoff values of CFI, TLI, and RMSEA perform with binary data generated with both compensatory and noncompensatory IRT models. We conclude this paper with conclusions and discussions, as well as some advice for applied researchers and practitioners who are interested in using model fit indices for unidimensionality assessment.

2. Literature Review

2.1 Simulation Studies on Model Fit Indices

In their highly influential study, Hu and Bentler[26] generated continuous data based on two model types (complex and simple), both of which assumed fifteen observed variables and three factors. They fixed the factor variances to 1.0 and the correlation between factors to 0.5, 0.4, and 0.3. For the simple model type, five variables load on each factor and there are no cross loadings; for the complex model type, one out of five variables that loads on one specific factor has a cross loading with another factor. They created seven data generation conditions by manipulating factors such as normality and correlation between factors and errors. To create scenarios of model misspecification, they either constrained the between factor correlation or some cross factor loadings to be zero. They created 200 replicated datasets within each condition and established the following cutoff values for model fit indices based on their simulation results: RMSEA < 0.06, CFI > 0.95, TLI > 0.95.

As Yu[35] decisively pointed out, Hu and Bentler's simulation study was based on maximum likelihood (ML) estimation method with continuous data, which is not suitable for categorical data usually estimated with robust diagonally weighted least square (DWLS) estimation methods. Since DWLS and ML use different fit functions and hence the chi-square values are different, the behavior of chi-square-based model fit indices might be different across different estimation methods and the cutoff values proposed out of simulation studies using one estimation method should not be generalized to other methods. She followed a similar simulation study design as Hu and Bentler's but focused on categorical outcomes and WLSMV estimator, the robust DWLS estimation method implemented in Mplus. She found that SRMR is not a good mode fit index for binary outcomes. TLI > 0.95 seems to perform satisfactorily when the sample size is greater than 250; for CFI, she found that CFI > 0.96 seems to perform better than CFI > 0.95; with RMSEA, she found that RMSEA < 0.05 outperforms RMSEA < 0.06.

Driven by the realization that Hu and Bentler's study was based on ML estimator with continuous outcomes and consequently, their proposed cutoff values of model fit indices might not generalize to cases of categorical outcomes estimated with DWLS estimator, Nye and Drasgow [34] conducted a simulation study to investigate how the cutoff values proposed by Hu and Bentler performed with binary data estimated with the DWLS estimator imple-

 DOI: https://doi.org/10.30564/jiep.v1i1.356

mented in LISREL 8.71[47]. Specifically, they simulated data from a two-factor model (the between factor correlation was fixed at 0.3) with 15 variables that either load on one factor or both factors; they manipulated the sample size to have three levels (400, 800, and 1600); they further manipulated the underlying distribution to have three levels (multivariate normal, moderately skewed, and severely skewed); they created misspecification scenarios by either constraining some factor loadings to be zero or both some factor loadings and between factor correlation to zero. They found that the commonly used cutoff values did not have enough power with DWLS estimator and more stringent values need to be used, and they concluded that simple cutoff values for model fit indices would not work since model fit can only be evaluated effectively in combination with the specific data.

While both Yu[35] and Nye and Drasgow[34] investigated the performances of these model fit indices with DWLS estimator and binary outcomes, neither included guessing in their data generation process and as a result, the effect of guessing upon the performance of model fit indices remains unclear. In addition, both simulations studies simulated data base on some factor analysis models and consequently, it is unknown how these model fit indices would perform with data generated with noncompensatory IRT models. Another difference is the nature of model misspecification: in both studies model misspecification takes the form of erroneous between-factor correlations or factor loadings, which is different from imposing a unidimensional structure upon data with multidimensional nature-the misspecification scenario we focus on in the current study.

2.2 Guessing in the Factor Analysis Framework

The effect of guessing is rarely investigated in the factor analysis framework. Among the few studies available, Carrol[48] found that when guessing was modeled, the tetrachoric correlations were corrected and hence stronger relation among the indicators were expected. In other words, if guessing was not modeled, the tetrachoric correlations would be attenuated relative to the true values. Considering that DWLS estimator is based on the estimation of tetrachoric correlations, such attenuation effects are expected to exist with data generated with guessing in the factor analysis framework.

Subsequent studies corroborate Carrol's findings. Tate[39] conducted a simulation study to investigate how guess affects decisions regarding dimensionality and parameter recovery in both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). He found that with guessing parameter fixed to be 0.2 in various multidimensional models, both EFA and CFA based on tetrachoric correlations uncorrected for guessing resulted in lower power to identify the true dimensionality. In terms of parameter recovery, the item thresholds and factor loadings in both EFA and CFA showed downward biases, which range in magnitude from 0.1 to 0.8 for item thresholds, and from 0.1 to 0.5 for factor loadings. Such biases were exacerbated with extreme item difficulties and discriminations.

Stone and Yeh[38] also investigated the guessing effect in EFA implemented in TESTFACT[49] using the Multistate Bar Examination data. They found that when guessing was modeled, the first eigenvalue of exploratory factor analysis (EFA) become larger and more items loaded substantially on factors. In addition, the average tetrachoric correlation increased from 0.07 to 0.11. Yeh conducted a large-scale simulation study to investigate guessing effect in EFA implemented in Mplus and TESTFACT[40]. Using a fixed sample size of 2,000 and a test length of 60 items, she systematically manipulated the number of dimensions, item discrimination parameters, between dimension correlations, and guessing magnitude to create various simulation conditions. Within each condition, 100 datasets were generated and estimated with EFA procedures implemented in both Mplus and TESTFACT. She found that TESTFACT, which allows the users to provide guessing values, outperformed Mplus in most simulation conditions regarding the ability to confirm the correct dimensionality.

2.3 Noncompensatory MIRT Model

Multidimensional IRT (MIRT)[50] models consist of compensatory and noncompensatory cases. Whereas the compensatory MIRT model is mathematically equivalent to a multi-factor model with categorical indicators (which is also known as a nonlinear factor model), the noncompensatory MIRT model does not have an equivalent counterpart in the factor analysis framework. The mathematical equation for a three parameter logistic noncompensatory MIRT model[51] takes the following form:

$$P(U_{ij}=1\,|\,\mathbf{è_i},\mathbf{a_j},\mathbf{b_j},c_j)=c_j+(1-c_j)\prod_{d=1}^{D}\frac{1}{1+\exp(-a_{jd}(\theta_{id}-b_{jd}))} \quad (1)$$

where U_{ij} is the response of examinee i to item j, D is the number of dimensions, θ_{id} is the ability of examinee i on dimension d, a_{jd} and b_{jd} are the discrimination parameter and difficulty parameter of item j on dimension d, and c_j is the lower asymptote of item j . As indicated by the Pi notation, the noncompensatory MIRT model assumes that inadequacy in one dimension cannot be completely compensated by adequacy in another dimension.

Comparing to a large number of methodological studies investigating the performance of various dimensionality assessment techniques with data generated using the compensatory model, there are considerably fewer

DOI: https://doi.org/10.30564/jiep.v1i1.356

ones focusing on data with noncompensatory structures. Among those few, Hattie, Krakowski, Rogers, and Swaminathan[42] investigated the performance of Stout's index of essential unidimensionality implemented in the DIMSEST software[5] with data generated using both the compensatory and noncompensatory model. They found that the DIMTEST procedure performed poorly when the data was generated with the latter model, and attributed its poor performance to the problems in estimating tetrachoric correlations. A more recent study is a simulation study conducted by Svetina[43], in which she generated data using the 2PL noncompensatory MIRT model and compared the performance of two methods (exploratory vs. cross validated) based on DETECT (Dimensionality Evaluation To Enumerate Contributing Traits)[52][18][53] and three methods[54][55][39] based on NOHARM (Normal Ogive Harmonic Analysis Robust Method)[56]. Having found that the performances of those methods can only be considered acceptable in a small number of conditions, she suggested that further studies be conducted before consideration of applying those methods to data suspected of having noncompensatory structure. It should be noted that in the above two studies, the magnitude of guessing was not systematically investigated: in the first study Hattie and et al. manipulated the lower asymptote to be either 0 or 0.15; Svetina only considered a 2PL model in which the guessing is assumed not to exist.

3. Methodology

3.1 Outcome Variable

A one-factor model was fit to each generated data set using Mplus with WLSMV estimator and RMSEA, CFI, and TLI were computed. For each of these three model fit indices, we computed its empirical power rate, which is the number of times that the model fit is considered poor divided by the number of replications within a simulation condition, using the following cutoff values: RMSEA > 0.05, CFI < 0.95, TLI < 0.95. It should be noted that for the sake of simplicity, in the following sections we use terms such as the power of RMSEA to refer to the power of using the cutoff value RMSEA > 0.05, the power of CFI for the power of using the cutoff value CFI < 0.95, and the power of CFI for the power of using the cutoff value TLI < 0.95.

3.2 Study Design

In both the compensatory and noncompensatory cases, the number of dimensions were fixed to three and following factors were manipulated:

1) Sample size (500, 1,000, or 2,000)
2) Number of items (30 or 60)
3) Between dimension correlation (0, 0.3, 0.5, and 0.7)
4) Pseudo-guessing value (0, 0.1, 0.2, 0.3, and 0.4)

For both the compensatory and noncompensatory cases, we have a fully crossed design with 3*2*4*5=120 conditions. Within each condition, we generated 1,000 datasets based on the corresponding MIRT model.

3.3 Item Response Generation

For both compensatory and noncompensatory cases, latent abilities were generated from three-dimensional multivariate normal distributions with a mean vector of 0s and a variance vector of 1s, and different levels of between dimension correlation values as specified in the previous section. Another commonality between the compensatory and noncompensatory cases is the systematic manipulation of pseudo-guessing values. The two cases differ regarding the generated item discrimination and difficulty parameter values.

For the compensatory case, the following three-dimensional three-parameter logistic (3PL) item response theory (IRT) model was used to generate item responses:

$$P(U_{ij}=1\,|\,\boldsymbol{\theta}_i,\mathbf{a}_j,d_j,c_j)=c_j+(1-c_j)\frac{1}{1+\exp(-\mathbf{a}_j\boldsymbol{\theta}_i+d_j))} \qquad (2)$$

where a_j is a vector of item *j*'s discrimination parameters on the three dimensions, θ_i is a vector of examinee i's scores on the three dimensions, d_j is item *j*'s multidimensional difficulty parameter, and c_j is the pseudo-guessing parameter. For item discrimination and difficulty parameters a_j and d_j, we used values provided by Reckase[50] as a realistic approximation of tests of simple structure. Since there are only 30 sets of item parameters, we generated another 30 similar items: for a_j, we added a random value drawn from N (0, 0.02) to each of the original 30 sets of discrimination parameters; for d_j, we added a random value drawn from N (0, 0.1) to each of the original 30 difficulty parameters. When the number of items is 30, only the first 30 sets of item discrimination and difficulty parameters were used for item response generation; when the number of items is 60, all 60 sets of item discrimination and difficulty parameters were used.

For the noncompensatory case, we followed the same item generating scheme adopted by Svetina[43]: item difficulty parameters were generated to fall in the range of -1.5 to 1.5 with an increment of 0.75, and item discrimination parameters on the dominant dimension were generated to range from 0.8 to 1.6 with an increment of 0.2, while on the remaining two dimensions they were generated to be 0.2 smaller than their counterparts on the dominant dimension. Items responses were generated based on the model specified in Equation 1.

 DOI: https://doi.org/10.30564/jiep.v1i1.356

4. Results

4.1 Compensatory Model

Table 1 lists the power rates of RMSEA to correctly reject unidimensionality across different simulation conditions; those of CFI and TLI appear in Tables 2-3. Specifically, the value within a cell indicates the number of times to reject unidimensionality divided by 1,000, when using the cutoff value for a given model fit index for data generated under a certain simulation condition. For example, the value 0.022 on the second row of Table 1 means that when sample size was fixed to 500 students and test length to 30 items, by applying RMSEA > 0.05 only 22 datasets were correctly identified as multidimensional out of the 1,000 datasets generated based on the 3PL MIRT model in Equation 2 with the between-dimension correlation equal to 0.5 (denoted as C3 in the table) and the pseudo-guessing parameter equal to 0.1 (denoted as G2). Similarly, the value 0.033 on the third row of Table 2 means that when sample size was fixed to 2,000 students and test length to 30 items, by applying CFI < 0.95 only 33 datasets were correctly identified as multidimensional out of the 1,000 datasets generated based on a 3PL MIRT model with the between-dimension correlation equal to 0.7 (denoted as C4 in the table) and the pseudo-guessing parameter equal to 0.2 (denoted as G3). As can be seen, a common pattern for the three model fit indices is that their power decreases with the decrease of sample size and the increase of guessing magnitude, between-dimension correlation, and test length.

Table 1. Power of RMSEA to Reject Unidimensionality in Compensatory Models

	SS = 500				SS = 1,000				SS = 2,000			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
30 Items												
G1	1	1	0.945	0	1	1	1	0	1	1	1	0
G2	1	0.986	0.022	0	1	1	0.015	0	1	1	0.049	0
G3	1	0.051	0	0	1	0.099	0	0	1	0.306	0	0
G4	0.284	0	0	0	0.707	0	0	0	0.949	0	0	0
G5	0	0	0	0	0	0	0	0	0	0	0	0
60 Items												
G1	1	1	0.013	0	1	1	0.198	0	1	1	0.941	0
G2	1	0.022	0	0	1	0.491	0	0	1	1	0	0
G3	0.305	0	0	0	1	0	0	0	1	0	0	0
G4	0	0	0	0	0	0	0	0	0	0	0	0
G5	0	0	0	0	0	0	0	0	0	0	0	0

Note. C1-C4 represent the between-dimension correlation value (C1=0, C2=0.3, C3=0.5, C4=0.7); G1-G5 refer to the pseudo-guessing parameter value (G1=0, G2=0.1, G3=0.2, G4=0.3, G5=0.5).

Table 2. Power of CFI to Reject Unidimensionality in Compensatory Models

	SS = 500				SS = 1,000				SS = 2,000			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
30 Items												
G1	1	1	0.986	0.04	1	1	1	0.022	1	1	1	0.023
G2	1	1	0.911	0.03	1	1	0.983	0.03	1	1	1	0.029
G3	1	1	0.669	0.036	1	1	0.805	0.028	1	1	0.985	0.033
G4	1	1	0.472	0.034	1	1	0.467	0.025	1	1	0.676	0.028
G5	1	1	0.476	0.028	1	1	0.286	0.026	1	1	0.283	0.031
60 Items												
G1	1	1	0.955	0.025	1	1	1	0.03	1	1	1	0.024
G2	1	1	0.678	0.029	1	1	0.982	0.026	1	1	1	0.024
G3	1	1	0.281	0.033	1	1	0.684	0.027	1	1	0.983	0.024
G4	1	1	0.206	0.025	1	1	0.194	0.027	1	1	0.295	0.021
G5	1	1	0.114	0.024	1	1	0.061	0.024	1	1	0.024	0.024

Note. C1-C4 represent the between-dimension correlation value (C1=0, C2=0.3, C3=0.5, C4=0.7); G1-G5 refer to the pseudo-guessing parameter value (G1=0, G2=0.1, G3=0.2, G4=0.3, G5=0.5).

Table 3. Power of TLI to Reject Unidimensionality in Compensatory Models

	SS = 500				SS = 1,000				SS = 2,000			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
30 Items												
G1	1	1	1	0.04	1	1	1	0.022	1	1	1	0.023
G2	1	1	0.964	0.03	1	1	1	0.03	1	1	1	0.029
G3	1	1	0.797	0.036	1	1	0.961	0.028	1	1	1	0.033
G4	1	1	0.472	0.034	1	1	0.691	0.025	1	1	0.798	0.028
G5	1	1	0.476	0.024	1	1	0.458	0.026	1	1	0.472	0.031
60 Items												
G1	1	1	0.955	0.025	1	1	1	0.03	1	1	1	0.024
G2	1	1	0.784	0.029	1	1	1	0.026	1	1	1	0.024
G3	1	1	0.477	0.033	1	1	0.795	0.027	1	1	1	0.024
G4	1	1	0.206	0.025	1	1	0.298	0.027	1	1	0.492	0.021
G5	1	1	0.186	0.024	1	1	0.061	0.024	1	1	0.062	0.024

Note. C1-C4 represent the between-dimension correlation value (C1=0, C2=0.3, C3=0.5, C4=0.7); G1-G5 refer to the pseudo-guessing parameter value (G1=0, G2=0.1, G3=0.2, G4=0.3, G5=0.5).

4.1.1 Without Guessing

Figure 1 shows how RMSEA, CFA, and TLI perform with the baseline conditions (no guessing) at various combinations of sample size, between-dimension correlation, and test length. Same as in tables 1-3, C1, C2, C3, and C4 on the horizontal axis represent the magnitude of between-dimension correlation; the values on the vertical axis, which range from zero to one, represent the statistical power of

DOI: https://doi.org/10.30564/jiep.v1i1.356

applying the cutoff value of a given model fit index under different simulation conditions. For example, the upper left panel displays the statistical power of applying RMSEA > 0.05 with data generated with thirty items, three sample sizes, four levels of between-dimension correlation, and without guessing.

As can be seen, when the correlation between the three dimensions is no greater than 0.5, all three indices have satisfactory power (greater than 0.9) to detect multidimensionality regardless of sample size and test length, with the exception of RMSEA having low power (lower than 0.2) when the sample size is either 500 or 1000 and the test length is 60 items. When the between-dimension correlation increases to 0.7, RMSEA, CFI, and TLI have extremely low power (lower than 0.1) regardless of sample size and test length. Sample size seems to have no considerable effect when the correlation between dimensions is no greater than 0.3, and the power only increases marginally with sample size increase when the correlation between dimensions is 0.5. One interesting pattern is that when the between-dimension correlation is 0.5, the performance of RMSEA is inversely related to the test length: its power with sample sizes of 500 and 1000 drops precipitously when the test length increases from 30 items to 60 items.

4.1.2 Impact of Guessing

As can be seen in Table 1, the power of RMSEA decreases with the increase of guessing magnitude, correlation between dimensions, and test length. Its power increases marginally with the increase of sample size. Compared with RMSEA, neither CFI nor TLI is influenced by the increase of guessing magnitude when the correlation between dimensions is low: both have a power of one when the correlation is 0.3 or lower. When the correlation is 0.5 or higher, however, the power of both decreases considerably with the increase of guessing magnitude regardless of the sample size and test length.

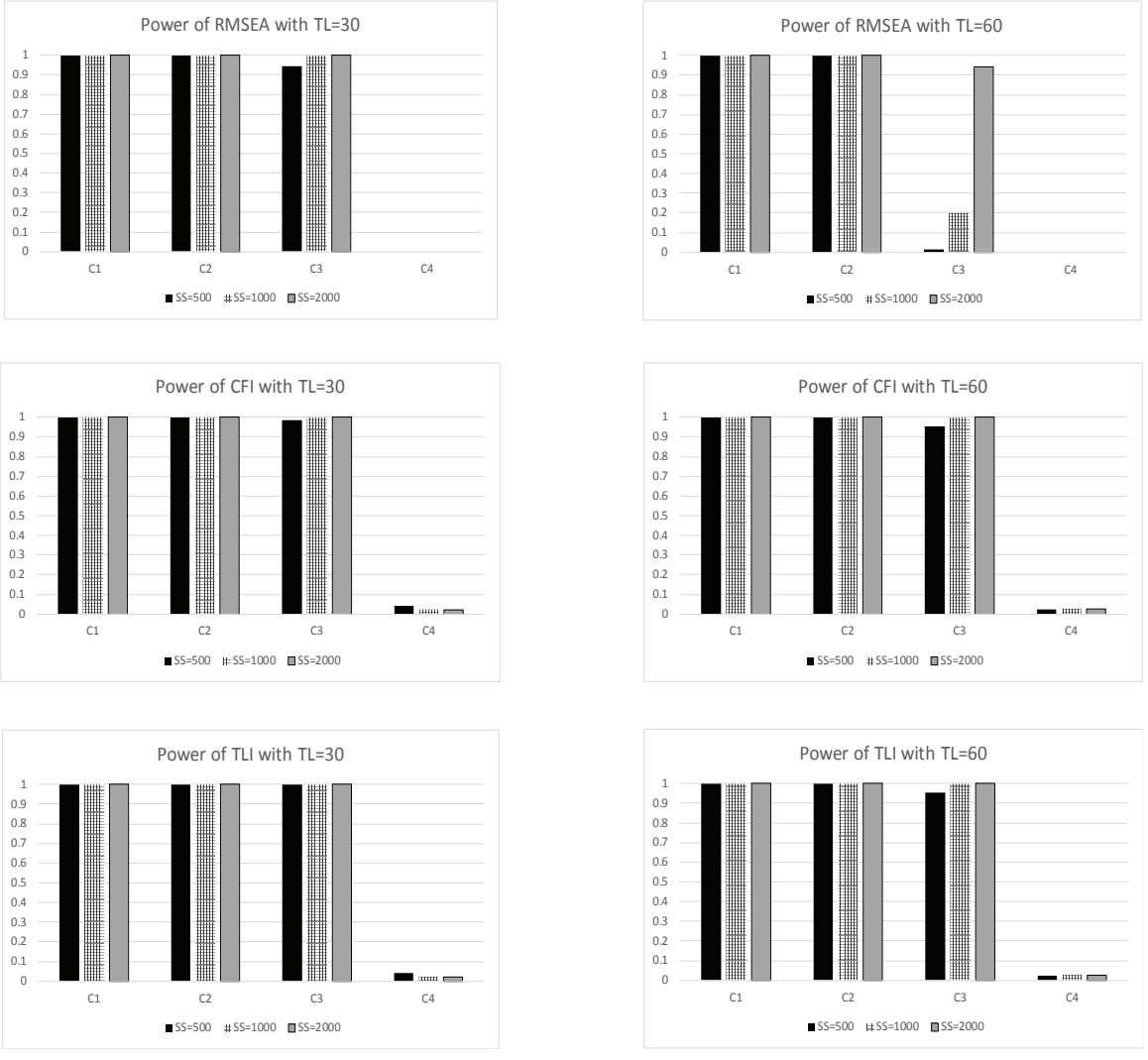

Figure 1. Power of RMSEA, CFI, and TLI with no guessing in compensatory models

 DOI: https://doi.org/10.30564/jiep.v1i1.356

To further explore the guessing impact upon the statistical power of RMSEA, CFI, and TLI, in Figure 2 we plot the average power rate of each model fit index across three sample sizes. Same as in tables 1-3, G1, G2, G3, and G4 on the horizontal axis represent the guessing magnitude, and the vertical axis represents the average statistical power across three sample sizes of applying the cutoff value of a given model fit index. For example, the bottom right panel displays the average statistical power of applying TLI < 0.95 across three sample sizes with data generated with sixty items, four levels of between-dimension correlation, and four levels of guessing magnitude.

For RMSEA, when the guessing is no greater than 0.1, the statistical power is close to one with the between-dimension-correlation no greater than 0.3 and the test length equal to 30; if the test length increases to 60, however, the power of RMSEA drops to 0.5 when the between-dimension-correlation is 0.3. When the guessing is 0.2, the statistical power is one only with the between-dimension-correlation is zero and the test length is 30; if the test length increases to 60, however, the power of RMSEA drops to slightly lower than 0.8. When the guessing is 0.3 or 0.4, RMSEA has no satisfactory statistical power regardless of the between-dimension-correlation and test length.

CFI and TLI seem to be robust to the guessing when the between-dimension correlation is no greater than 0.3: their power remains invariably close to one regardless of the guessing value and sample size. When the between-dimension correlation is 0.5, guessing seems to have a systematic influence: the power of both CFI and TLI decreases with the increase of guessing value. Test length also plays a role regarding the power of CFI and TLI when the between-dimension correlation is 0.5: both indices have consistently lower power when the test length is 60 items than when it is 30 items. When the test length is 30 items and the between-dimension correlation is 0.5, both CFI and TLI have satisfactory power (greater than 0.8) when guessing is no greater than 0.2; when the test length is 60 items and the between-dimension correlation is 0.5, CFI

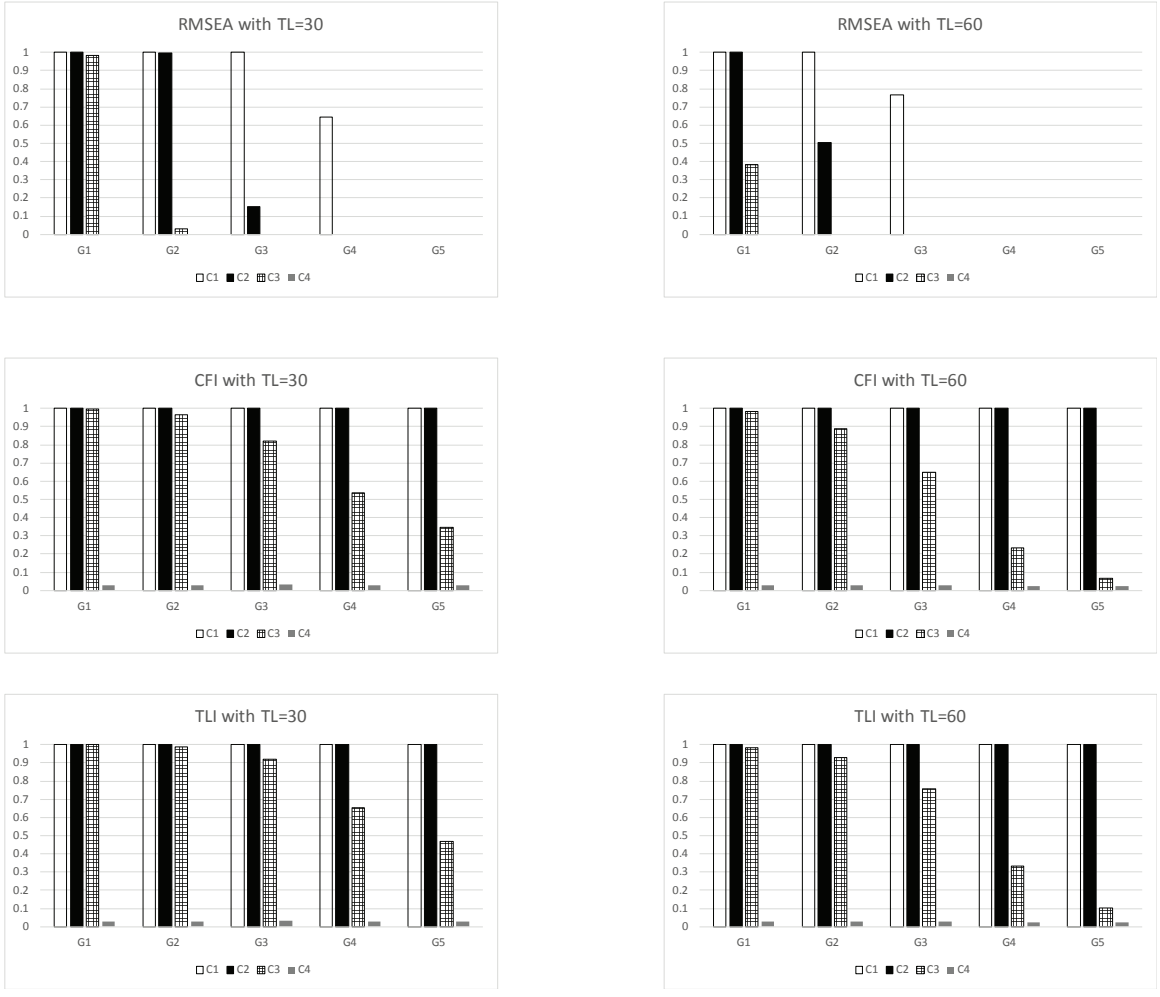

Figure 2. Power of RMSEA, CFI, and TLI with different guessing levels in compensatory models

DOI: https://doi.org/10.30564/jiep.v1i1.356

and TLI have satisfactory power (greater than 0.8) only when guessing is no greater than 0.1. Similar to RMSEA, neither CFI nor TLI has enough statistical power when the between-dimension-correlation is 0.7 regardless of guessing value, sample size, and test length.

4.2 Noncompensatory Model

As the power of RMSEA to reject unidimensionality are invariably zero across all simulation conditions, we focus on the power of CFI and TLI in the concompensatory cases. Table 4 lists the power of CFI to reject unidimensionality across different simulation conditions, and those of TLI appear in Table 5. Different than the pattern observed in the preceding compensatory model that guessing magnitude, between-dimension correlation, sample size, and test length systematically affect the power of RMSEA, CFA, and TLI, here the only discernable pattern is that the power of these three model fit indices decreases with the increase of between-dimension correlation. In terms of guessing, although the power of RMSEA, CFA, and TLI change with the change of guessing magnitude, the change is not in a systematic pattern as in the compensatory cases. Another difference is that the power of RMSEA, CFA, and TLI observed here seem to be considerably lower than in the compensatory model.

4.2.1 Without Guessing

As RMSEA has no statistical power regardless of the guessing value, the between-dimension-correlation, and test length, we focus on CFI and TLI regarding their performances with the baseline condition (no guessing). Figure 3 plots the mean power rates of these two model fit indices across averaged across four between-dimension correlation values. Regardless of the test length, their statistical power becomes satisfactory (CFI has statistical power slightly lower than 0.8 when the test length is 30) only when the between-dimension-correlation value is zero. In contrast to what has been observed in the compensatory cases where statistical power of RMSEA, CFA, and TLI decreases with the increase of test length, with zero between-dimension correlation CFI and TLI have slightly higher statistical power when the test length is 60 items than when it is 30 items.

4.2.2 Impact of Guessing

As the power of RMSEA remains zero in all simulation conditions, it is not possible to evaluate the effect of guessing upon the performance of RMSEA. In this section we focus on how guessing affects the performances of CFI and TLI in the noncompensatory cases. As can be seen from Tables 4-5, when data were generated with a noncompensatory IRT model, while the power of CFA and TLI seem to decrease with the increase of between-dimension correlation value, there seems to be no discernable patterns regarding how their power change as a result of the change of guessing value, test length, or sample size.
To further explore the guessing impacts upon the statistical power of these two model fit indices, in Figure 4 we

Table 4. Power of CFI to Reject Unidimensionality in Noncompensatory Models

	SS = 500				SS = 1,000				SS = 2,000			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
30 Items												
G1	0.789	0.104	0	0	0.451	0	0	0	0.085	0	0	0
G2	0.482	0.297	0.032	0.054	0.292	0.024	0	0	0.206	0	0	0
G3	0.421	0.275	0.06	0.108	0.438	0.123	0	0	0.21	0	0	0
G4	0.414	0.407	0.194	0.205	0.427	0.203	0.076	0.079	0.27	0.068	0	0
G5	0.411	0.406	0.258	0.266	0.387	0.264	0.19	0.221	0.28	0.184	0.088	0
60 Items												
G1	0.989	0	0	0	0.502	0	0	0	0.189	0	0	0
G2	0.818	0.052	0.03	0	0.806	0.062	0	0	0.804	0.017	0	0
G3	0.677	0.21	0.055	0	0.485	0.051	0	0	0.488	0	0	0
G4	0.651	0.299	0.212	0.061	0.482	0.123	0	0	0.492	0.008	0	0
G5	0.619	0.462	0.279	0.225	0.452	0.296	0.071	0.017	0.475	0.111	0	0

Note. C1-C4 represent the between-dimension correlation value (C1=0, C2=0.3, C3=0.5, C4=0.7); G1-G5 refer to the pseudo-guessing parameter value (G1=0, G2=0.1, G3=0.2, G4=0.3, G5=0.5).

Table 5. Power of TLI to Reject Unidimensionality in Noncompensatory Models

	SS = 500				SS = 1,000				SS = 2,000			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
30 Items												
G1	0.921	0.197	0	0	0.451	0	0	0	0.175	0	0	0
G2	0.513	0.297	0.063	0.054	0.479	0.059	0	0.028	0.297	0	0	0
G3	0.512	0.283	0.1	0.1	0.497	0.112	0.011	0.017	0.292	0.014	0	0
G4	0.53	0.515	0.186	0.201	0.539	0.194	0.099	0.06	0.295	0.051	0	0
G5	0.651	0.543	0.284	0.295	0.56	0.292	0.187	0.219	0.518	0.184	0.061	0
60 Items												
G1	0.989	0	0	0	0.502	0	0	0	0.293	0	0	0
G2	0.819	0.099	0.03	0	0.806	0.106	0	0	0.908	0.027	0	0
G3	0.695	0.206	0.053	0	0.493	0.051	0	0	0.488	0	0	0
G4	0.709	0.493	0.203	0.048	0.522	0.193	0	0	0.499	0.016	0	0
G5	0.764	0.519	0.286	0.322	0.531	0.301	0.104	0.012	0.517	0.1	0	0

Note. C1-C4 represent the between-dimension correlation value (C1=0, C2=0.3, C3=0.5, C4=0.7); G1-G5 refer to the pseudo-guessing parameter value (G1=0, G2=0.1, G3=0.2, G4=0.3, G5=0.5).

DOI: https://doi.org/10.30564/jiep.v1i1.356

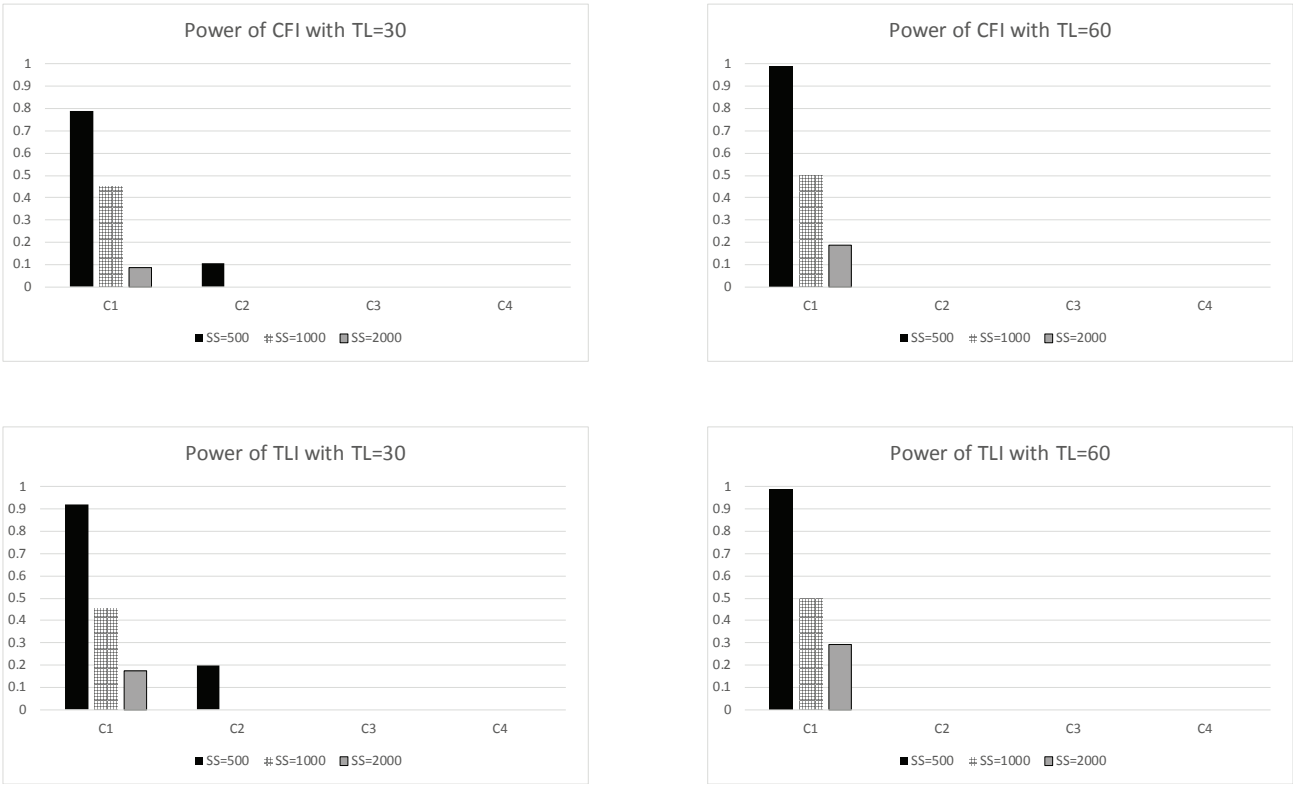

Figure 3. Power of RMSEA, CFI, and TLI with no guessing in noncompensatory models

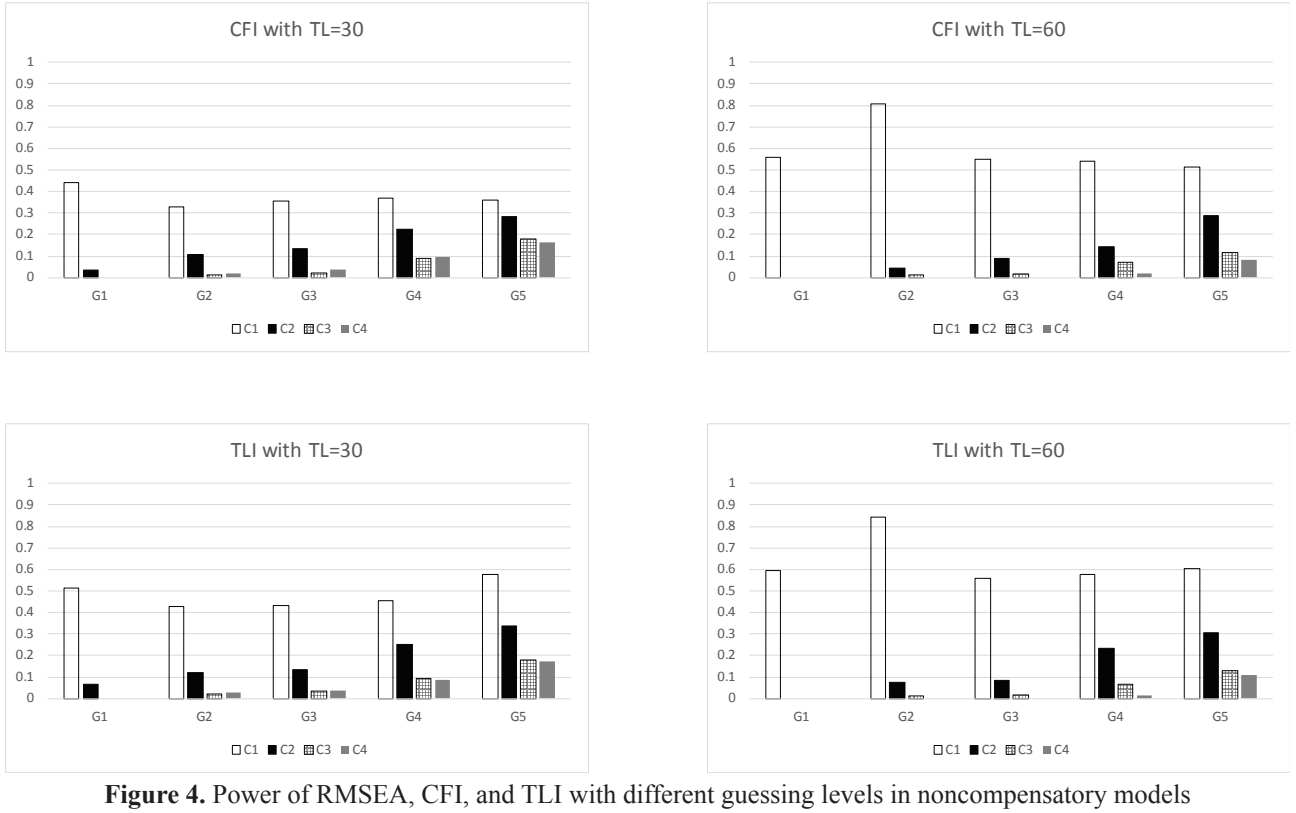

Figure 4. Power of RMSEA, CFI, and TLI with different guessing levels in noncompensatory models

DOI: https://doi.org/10.30564/jiep.v1i1.356

plot the average power rate of CFI and TLI across three sample sizes. As can be seen, CFI and TLI have power greater than 0.8 only when between-dimension correlation is zero and the guessing value is 0.1; however, as far as guessing is concerned, despite the fact that the power of CFI and TLI changes with the change of guessing magnitude, no patterns can be observed that suggest a systematic influence of guessing upon their statistical power to detect multidimensionality.

5. Discussion and Conclusion

It should be noted that due to various warnings and caveats against the use of cutoff values for model fit indices for model fit assessment, methodologists have developed an equivalence testing approach[57][58] that does not rely on common cutoff values but create adjusted cutoff values. Such methodological developments notwithstanding, the use of cutoff values for model fit indices to assess model fit still remains hugely popular among studies across many disciplines[59][60][61][62][63][64].

The purpose of this study was to reiterate the important point that cut off values for model fit indices should never be used indiscriminately for dimensionality assessment. In addition, we explored how guessing and the nature of multidimensionalty, two factors ignored in previous studies, could further negatively affect the performances of cutoff values for model fit indices in dimensionality assessment. Specifically, we systematically investigated how guessing affected the statistical power of commonly used cutoff values for RMSEA, CFI, and TLI to refute unidimensionality with binary data generated with either compensatory or noncompensatory IRT models. It was hypothesized that as all the simulation studies which established the commonly used cutoff values for model fit indices were based on factor analysis models, which do not accommodate guessing, such cutoff values (RMSEA < 0.05; CFI > 0.95; TLI > 0.95) would exhibit poor statistical power with binary data generated with IRT models that include a guessing parameter within.

The simulation results show that when data were generated with a 3PL compensatory multidimensional IRT model, increases of guessing value lead to decreases of the power of RMSEA, and such decreases were exacerbated with the increase of between-dimension correlation. For CFA and TLI, when the between-dimension correlation was no greater than 0.3, they were robust to guessing effect and their power remained constantly one regardless of the guessing magnitude and sample size; the systematic effect of guessing upon the power of CFA and TLI appeared when the between-dimension correlation was 0.5, in that their power decreased with the increase of guessing magnitude, and such decreases became more pronounced with a longer test length. When the between-dimension correlation was 0.7, all three indices had virtually no power to detect multidimensionality. When data were generated with a 3PL noncompensatory multidimensional IRT model, guessing did not have a systematic effect upon the statistical power of the three model fit indices, although it should be noted that a small change of guessing magnitude can result in a considerable change of statistical power for a given model fit index. For example, as can be observed in Table 6, when the sample size was 500 and the between-dimension correlation was zero, the power of TLI dropped from 0.921 to 0.513 when the guessing magnitude changed from zero to 0.1.

We also investigated how the cutoff values performed with the baseline conditions in which the guessing value was zero (the model reduced to a 2PL compensatory/noncompensatory IRT model). In the compensatory case, it was found that when the between-dimension correlation was no greater than 0.5, CFI and TLI exhibited statistical power higher than 0.90 regardless of test length and sample size; RMSEA displayed the same pattern when the test length was 30 items. When the test length was 60 items, RMSEA performed poorly when the between-dimension correlation was 0.5 with sample size equal to 500 or 1000, and its statistical power went up to 0.941 when the sample size was 2000. None of the model fit indices performed satisfactorily when the between-dimension-correlation was 0.7, regardless of sample size and test length. It seems that when such high correlations exist between dimensions, none of RMSEA, CFA, and TLI can statistically differentiate such structures from unidimensional structure. In the noncompensatory case, it was found that RMSEA had no power at all to detect multidimensionality regardless of the sample size, test length, and between-dimension correlation. CFI and TLI displayed unsatisfactory power (less than 0.8) in most conditions with some exceptions: CFI had a power of 0.989 when the sample size was 500, the test length was 60 items, and the between-dimension correlation was zero; TLI had power greater than 0.9 when the test length was 60 items, and the between-dimension correlation was zero.

The well-known advice that model fit indices should not be used indiscriminately is corroborated by the results found in the baseline conditions where no guessing is assumed to exist. Apparently, the power of CFI, TLI, and RMSEA is affected by the test length in that (a) with the same sample size, a longer test results in decreased power of the three model fit indices, and (b) a larger sample size is required for the three model fit indices to perform well in a longer test. Taking the perspective that the degree

 DOI: https://doi.org/10.30564/jiep.v1i1.356

to which a model is misspecified is determined by the statistical power to detect such misspecifications[30], we conclude that with the same generating model and same model misspecification type, increased test lengths result in amelioration of model misspecification due to the reduced statistical power: it is more difficult to detect model specification of a less misspecified model. Although we only investigated two test lengths in the current study, it is expected that with tests consist of more than 60 items, the power of these three model fit indices will be lower than those presented in Tables 1-3. Another reason that the common cutoff values cannot be generalized is that the magnitude of factor loadings impact their performances. Heene, Hilbert, Draxler, and Ziegler[65] found that the statistical power of RMSEA, SRMR, and CFI changes with the change of the magnitude of factor loadings. In other words, if data were simulated with different item parameters, the results in Tables 1-3 might not be replicated. This is further evidence that "golden rules" are extremely difficult, if not possible, to find.

The findings that the performances of RMSEA, CFA, and TLI are subject to guessing effect are hardly surprising. When guessing effect exists, the measurement quality deteriorates, and as nicely stated by Hancock and Muller[37], "as measurement quality gets poorer, common data-model fit indices-absolute, parsimonious, and/or incremental in nature-paint an increasingly and deceptively favorable picture of the model's latent structure." In other words, the guessing effect introduces noise into data, which can mask the true latent structure. What is surprising, however, is that a small increase of guessing magnitude can result in precipitous decrease of the statistical power of a certain model fit index considered in this study. Take RMSEA as an example: as can be seen in Table 1, when guessing increases from 0 to 0.1, its power drops from 1 to 0.015 with a samples size of 1000 and a test length of 30 items. CFI and TLI do not have such drastic changes of power as RMSEA does, yet an increase of 0.1 of guessing magnitude can still result in a decrease of 0.2 to 0.3 regarding their statistical power.

One piece of advice to practitioners and researchers who are interested in using model fit indices to assess unidimensionality is that the consequent conclusions regarding unidimensionality should be taken with a grain of salt and interpreted cautiously, especially with binary data that represent scores on multiple-choice questions. As shown in this study, existence of guessing decreases the sensitivity of RMSEA, CFA, and TLI to multidimensionality. It is recommended that if model fit indices are used for unidimensionality assessment, other techniques such as DIMTEST and DETECT that can model guessing should be used jointly, although it is possible that different methods might disagree with each other[66]. When facing inconsistence dimensionality assessment results from different methods, we recommend using the bifactor modeling approach[16][67], which, unlike the other unidimensionality assessment approaches that attempt to provide a yes/no answer regarding unidimensionality, provides a detailed picture of the consequence of treating the data as unidimensional and allows one to empirically examine whether and how the model parameter estimates change by fitting a unidimensional structure to a multidimensional data set.

One limitation of the current study is that in the compen-satory cases, we generated data using item parameters that were designed to realistically mimic a simple structure of multidimensionality with each item predominantly measuring one dimension. Although not strictly a simple structure, the generating items are distinct from those used to mimic a complex structure of multidimensionality, and it is expected that the cutoff values of the three model fit indices considered in the current study will perform dif-ferently with items following a complex structure.

Taken together, the results in the present study show that when data follow a compensatory multidimensional structure, guessing systematically decreases the power of the commonly used cutoff values of RMSEA, CFA, and TLI. When data follow a noncompensatory multidimensional structure, these cutoff values do not perform well and guessing does not seem to affect their power in a systematic manner. Such findings point to two directions for possible future research. First, as guessing systematically affects the distribution of model fit indices, which is another reason cutoff values for these model fit indices should not be used, the performances of other more recent methods such as the equivalent testing approach mentioned earlier and the permutation test[68][69], when dealing with data containing guessing, should be investigated. The finding that cutoff values for model fit indices performed poorly in assessing the dimensionality of data generated with noncompensatory models, together with those by Hattie, Krakowski, Rogers, and Swaminathan[42] and Svetina[43], suggest that dimensionality assessment techniques that are based on the compensatory framework do not work well with noncompensatory data. In that regard, methods specifically designed for noncompensatory data are direly needed.

References

[1] Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. Applied Psychological Measurement, 13(2), 113-127.

  DOI: https://doi.org/10.30564/jiep.v1i1.356

[2] Dorans, N. J., & Kingston, N. M. (1985). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating of the GRE verbal scale. Journal of Educational Measurement, 22(4), 249-262.

[3] Kirisci, L., Hsu, T. C., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. Applied Psychological Measurement, 25(2), 146-162.

[4] Zhang, B. (2010). Assessing the accuracy and consistency of language proficiency classification under competing measurement models. Language Testing, 27(1), 119-140.

[5] Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. Psychometrika, 52(4), 589-617.

[6] Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. Educational and Psychological Measurement, 37(4), 827-838.

[7] Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. Applied Psychological Measurement, 9(2), 139-164.

[8] Raykov, T., & Pohl, S. (2013). Essential unidimensionality examination for multicomponent scales: an interrelationship decomposition approach. Educational and Psychological Measurement, 73(4), 581-600.

[9] Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. Psychometrika, 55(2), 293-325.

[10] Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. Psychometrika, 72(2), 141.

[11] Bejar, I. I. (1980). A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. Journal of Educational Measurement, 17(4), 283-296.

[12] Christensen, K. B., Bjorner, J. B., Kreiner, S., & Petersen, J. H. (2002). Testing unidimensionality in polytomous Rasch models. Psychometrika, 67(4), 563-574.

[13] Debelak, R., & Arendasy, M. (2012). An algorithm for testing unidimensionality and clustering items in Rasch measurement. Educational and Psychological Measurement, 72(3), 375-387.

[14] Heene, M., Kyngdon, A., & Sckopke, P. (2016). Detecting violations of unidimensionality by order-restricted inference methods. Frontiers in Applied Mathematics and Statistics, 2, 3.

[15] McDonald, R. P., & Mok, M. M. C. (1995). Goodness of fit in item response models. Multivariate Behavioral Research, 30(1), 23-40.

[16] Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of Life Research, 16(1), 19-31.

[17] Weng, L. J., & Cheng, C. P. (2005). Parallel analysis with unidimensional binary data. Educational and Psychological Measurement, 65(5), 697-716.

[18] Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. Psychometrika, 64(2), 213-249.

[19] Millsap R. E. (2011). Statistical Approaches to Measurement Invariance. Taylor and Francis Group: New York.

[20] Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

[21] Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. Structural Equation Modeling, 15(1), 136-153.

[22] Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. Multivariate Behavioral Research, 26(3), 457-477.

[23] Luo, Y. (2018). A short note on estimating the testlet model with different estimators in Mplus. Educational and Psychological Measurement, 78(3), 517-529.

[24] Luo, Y., & Dimitrov, D. M. (2018). A short note on obtaining point estimates of the IRT ability parameter with MCMC estimation in Mplus: how many plausible values are needed? Educational and Psychological Measurement.

[25] Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52(3), 393-408.

[26] Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1-55.

[27] Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. Structural Equation Modeling, 22(4), 504-516.

[28] Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. Structural Equation Modeling, 12(1), 41-75.

[29] Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. Structural Equation Modeling, 12(3), 343-367.

[30] Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. Multivariate Behavioral Research, 42(3), 509-529.

 DOI: https://doi.org/10.30564/jiep.v1i1.356

[31] Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Structural Equation Modeling, 11(3), 320-341.

[32] Yuan, K. H. (2005). Fit indices versus test statistics. Multivariate Behavioral Research, 40(1), 115-148.

[33] Huggins-Manley, A. C., & Han, H. (2016). Assessing the sensitivity of weighted least squares model fit indexes to local dependence in item response theory models. Structural Equation Modeling, 24(3), 331-340.

[34] Nye, C. D., & Drasgow, F. (2011). Assessing goodness of fit: Simple rules of thumb simply do not work. Organizational Research Methods, 14(3), 548-570.

[35] Yu, C. Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes (Doctoral dissertation, University of California Los Angeles).

[36] McNeish, D., An, J., & Hancock, G. R. (2017). The thorny relation between measurement quality and fit index cutoffs in latent variable models. Journal of Personality Assessment.

[37] Hancock, G. R., & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. Educational and Psychological Measurement, 71(2), 306-324.

[38] Stone, C. A., & Yeh, C. C. (2006). Assessing the dimensionality and factor Structure of multiple-choice exams an empirical comparison of methods using the multistate bar examination. Educational and Psychological Measurement, 66(2), 193-214.

[39] Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. Applied Psychological Measurement, 27(3), 159-203.

[40] Yeh, C. C. (2007). The effect of guessing on assessing dimensionality in multiple-choice tests: A Monte Carlo study with application (Doctoral dissertation, University of Pittsburgh).

[41] Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. Applied Psychological Measurement, 27(6), 395-414.

[42] Hattie, J., Krakowski, K., Jane Rogers, H., & Swaminathan, H. (1996). An assessment of Stout's index of essential unidimensionality. Applied Psychological Measurement, 20(1), 1-14.

[43] Svetina, D. (2013). Assessing dimensionality of noncompensatory multidimensional item response theory with complex structures. Educational and Psychological Measurement, 73(2), 312-338.

[44] Muthén, L., & Muthén, B. (1998-2012). Mplus User's Guide (Seventh Edition). Los Angeles, Ca: Muthén & Muthén.

[45] Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished manuscript.

[46] West, S. G., Taylor, A. B., & Wu, W. (in press). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), Handbook of Structural Equation Modeling. New York: Guilford Press.

[47] Joreskog, K., & Sorbom, D. (1996). User's reference guide. Chicago, IL: Scientific Software International.

[48] Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. Psychometrika, 10(1), 1-19.

[49] Wilson, D.,Wood, R., Gibbons, R., Schilling, S., Muraki, E., & Bock, R. D. (2003). TESTFACT: Test scoring and full information item factor analysis (Version 4.0). Lincoln-wood, IL: Scientific Software International.

[50] Reckase, M. (2009). Multidimensional item response theory (Vol. 150). New York: Springer.

[51] Sympson, J. B. (1978). A model for testing with multidimensional items. In D. J. Weiss (Ed.). Proceedings of the 1977 Computerized Adaptive Testing Conference (pp. 82-89). Minneapolis, MN: University of Minneapolis, Department of Psychology, Psychometric Methods Program.

[52] Kim, H. R. (1994). New techniques for the dimensionality assessment of standardized test data. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

[53] Zhang, J., & Stout, W. F.(1999). Conditional covariance structure of generalized compensatory multidimensional items. Psychometrika, 64(3), 129-152.

[54] Gessaroli, M. E., & De Champlain, A. F. (1996). Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. Journal of Educational Measurement, 33(2), 157-179.

[55] Gessaroli, M. E., De Champlain, A. F., & Folske. (1997, March). Assessing dimensionality using a likelihood-ratio chi-square test based on a non-linear factor analysis of item response data. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

[56] Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. Multivariate Behavioral Research, 23(2), 267-269.

[57] Yuan, K. H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indexes. Structural Equation Modeling: A Multidisciplinary Journal, 23(3), 319-330.

[58] Marcoulides, K. M., & Yuan, K. H. (2017). New ways to evaluate goodness of fit: A note on using equivalence test-

  DOI: https://doi.org/10.30564/jiep.v1i1.356

ing to assess structural equation models. Structural Equation Modeling: A Multidisciplinary Journal, 24(1), 148-153.

[59] Blömeke, S., Dunekacke, S., & Jenßen, L. (2017). Cognitive, educational and psychological determinants of prospective preschool teachers' beliefs. European Early Childhood Education Research Journal, 1-19.

[60] Campbell, P., Hope, K., & Dunn, K. M. (2017). The pain, depression, disability pathway in those with low back pain: a moderation analysis of health locus of control. Journal of Pain Research, 10, 2331-2339.

[61] Cougle, J. R., Summers, B. J., Allan, N. P., Dillon, K. H., Smith, H. L., Okey, S. A., & Harvey, A. M. (2017). Hostile interpretation training for individuals with alcohol use disorder and elevated trait anger: a controlled trial of a web-based intervention. Behaviour Research and Therapy, 99, 57-66.

[62] Drinkwater, K., Denovan, A., Dagnall, N., & Parker, A. (2017). An assessment of the dimensionality and factorial structure of the revised paranormal belief scale. Frontiers in Psychology, 8, 1693.

[63] Firmin, R. L., Lysaker, P. H., McGrew, J. H., Minor, K. S., Luther, L., & Salyers, M. P. (2017). The Stigma Resistance Scale: A multi-sample validation of a new instrument to assess mental illness stigma resistance. Psychiatry Research, 258, 37-43.

[64] Govender, K., Cowden, R. G., Asante, K. O., George, G., & Reardon, C. (2017). Validation of the child and youth resilience measure among South African adolescents. PloS one, 12(10), e0185815.

[65] Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: a cautionary note on the usefulness of cutoff values of fit indices. Psychological Methods, 16(3), 319.

[66] Luo, Y., & Al-Harbi, K. (2016). The utility of the bifactor method for unidimensionality assessment when other methods disagree: an empirical illustration. SAGE Open, 6(4), 2158244016674513.

[67] Reise, S. P. (2012). The rediscovery of bifactor measurement models. Multivariate Behavioral Research, 47(5), 667-696.

[68] Jorgensen, T. D., Kite, B. A., Chen, P. Y., & Short, S. D. (2017). Permutation randomization methods for testing measurement equivalence and detecting differential item functioning in multiple-group confirmatory factor analysis. Psychological Methods, 23(4), 708-728.

[69] Kite, B. A., Jorgensen, T. D., & Chen, P. Y. (2018). Random permutation testing Applied to measurement invariance testing with ordered-categorical indicators. Structural Equation Modeling: A Multidisciplinary Journal, 25(4), 573-587.

 DOI: https://doi.org/10.30564/jiep.v1i1.356

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep/

ARTICLE

Determinants of Vietnamese Colleges' Academic Performance: The Second Stage Bootstrapping DEA Approach

Carolyn-Dung Thi Thanh Tran*
International College of Management, Sydney, NSW 2095, Australia

ARTICLE INFO

Article history:
Received: 29 November 2018
Accepted: 21 December 2018
Published: 18 January 2019

Keywords:
Efficiency
Data envelopment analysis
Bootstrapping
Colleges
Vietnam

ABSTRACT

This paper aims to examine the determinants of the academic performance of 141 colleges in Vietnam in the period 2011/12–2013/14. The second-step bootstrapping data envelopment analysis is proposed to measure the performance of colleges and examine the influences of environmental variables on their operational efficiency. The results disclose that colleges are not technically efficient in their operations. To attain the frontier efficiency, colleges could potentially advance their performance, on average, 37.7%. The inefficiencies of colleges can be induced by external factors including location, age, and ownership presented as key influencers. Our findings are anticipated to offer more insights of the performance of colleges for policy makers on the way seeking possible solutions to enhancing innovation in performance of Vietnamese colleges.

1. Introduction

In the trends of globalization and internationalization today, institutions of higher education (HEIs) has been attracting much attention of policy makers because the human resource development of a nation substantially influences its economic growth.[6] However, due to challenges of fiscal policy of many nations, the distribution of public services is required for greater efficiency. Accordingly, the performance of HEIs would be asked to be more efficient.[3,35] This is to say that assessing the performance of HEIs plays a crucial role to address community's concern and increase transparent accountability to society. [3,35,4].

Vietnam started its economic reform from 1986, strongly linked to renovation in higher education. However, not untill 1997, when the government introduced a new policy (Resolution 90/1997/NQ-CP), in which private education

**Corresponding Author:*
Carolyn-Dung Thi Thanh Tran,
Discipline Leader in Business, International College of Management, Sydney, NSW 2095*; E-mail: ctran@icms.edu.au.*
Adjunct Lecturer, the University of New England Business School, Armidale, NSW 2351*, Australia; E-mail: ttran*43*@une.edu.au*

DOI: https://doi.org/10.30564/jiep.v1i1.380

was officially encouraged, Vietnamese higher education has made remarkable growth in enrolments (142%) and in numbers of universities and colleges (155%) for the period 2000/01–2016/17. [29] Government investment in education has kept stable at least 20% of total national expenditure, in which the average recurrent expenditure for education and training is 11% at the central level and 89% at the local level.[19] In addition, as in comparison with the national budget for education in 2013, the investment for education in 2017 augmented by 59%. [45] This implies the great efforts of the government to enhance the innovation of higher education.

Together with this, the government promulgated numerous policies to support operations of higher education such as Resolution 14/2005/NQ-CP for innovation in higher education (the Higher Education Reform Agenda) for 2006–2020 and Decision 711/2012/QD-TTg in terms of innovative strategy in higher education for the phase of 2011-2020. However, whether HEIs are working well under the currently legal environment and if external uncontrolled factors can affect their performance have been less concerned in empirical research in Vietnam.

Our main objective is to investigate operational efficiency of Vietnamese colleges and impacts of determinants that would cause colleges to be inefficient in their performance. Such evaluation is fitting and timely to understand better colleges' performance on the way to seek feasible solutions for innovation and creativity in academic operations. The structure of our study is planned as follows. Section 2 briefly discusses the main features of Vietnamese colleges. Section 3 presents a brief literature review on efficiency measurement in higher education. The method of analysis is presented in Section 4. Data sources and variables are introduced in Section 5. The empirical results of the study are presented in Section 6. Discussion and conclusions are provided in Section 7.

2. Vietnamese Colleges

Vietnamese HEIs comprise universities and colleges. According to the Ministry of Education and Training,[29] there are a total of 444 HEIs including 235 universities and 219 colleges. Although they are classified as HEIs according to the 2012 Education Law, they face heterogeneous teaching environments. Colleges train students with associate bachelor degrees (12+3). They do not offer the postgraduate programs and tend to focus on academic consultant services rather than academic research. Accordingly, different sets of regulations are applied to them due to their distinct states of nature. That is to say that government emphasises considerations at different levels on assessment of performance of universities and colleges.

Table 1. Growth of Colleges over the years

	2000/01	2005/06	2010/11	2016/17
Institutions				
Public	99	142	196	187
Private	5	12	30	30
Total	104	154	226	217
Enrolment				
Public	171,922	277,176	581,829	516,296
Private	14,801	22,118	144,390	76,897
Total	173,912	299,294	726,216	593,193
Share of Total National Enrolment				
Public	19%	20%	27%	20%
Private	2%	2%	7%	4%
Total	21%	22%	34%	24%

Sources: MOET (2018).

While policy makers place much concern on development of universities, the role and position of colleges have seemed to be ignored. In actual fact, colleges have contributed significantly to higher education and economy by supplying a highly qualified technician force to the labour market.

In 2000/01, enrolments of colleges accounted for 21% of total national enrolments. This figure increased to 34% in 2010/11. However, this figure witnessed a slight decrease in 2016/17 to 24%. Although the contribution of colleges to higher education is sound and worth being recognized, a reduction in total enrolments in recent years may be a problem to be considered. However, little research on performance of colleges has been undertaken to see what the performance of colleges is and whether there are any determinants that could potentially affect performance of colleges, causing them to be inefficient in the process of academic operations towards innovation and creativity in learning and teaching activities. Hence, it is timely to investigate these determinants for the aim of improving performance of colleges.

3. Efficiency of Higher Education: A Review

Measuring the operational efficiencies of higher education has been the focus of many empirical studies in recent years. Efficiency refers to comparing the current performance to the best practice and measured by setting outputs of a production unit against the inputs it uses.[22] Put differently, efficiency technically measures the ability of individual decision-making units (DMUs) to produce maximum output given the levels of inputs or produce the given levels of outputs at the minimum levels of the

 DOI: https://doi.org/10.30564/jiep.v1i1.380

inputs. The two most common methods for estimating the performance of HEIs are currently data envelopment analysis (DEA) and stochastic frontier analysis, which have been widely applied for different organisations across various countries. However, DEA is more well-known to scholars and researchers since multiple inputs and outputs can be used simultaneously without price information. Moreover, prior assumption of the relationship between inputs and outputs is not required as well. In this sense, we aim to provide a brief literature review on efficiency measurement in higher education using DEA that is relevant and applicable to Vietnamese colleges' context.

Before 2000, the majority of studies focused on assessing efficiencies of HEIs in developed countries such as the US, the UK, Australia, and Canada. Ahn, Charnes and Cooper[7] estimated the technical and scale efficiency for 161 US HEIs that were grouped according to whether or not they had a medical school in 1984/85. Their findings indicate that private institutions are less efficient than public institutions excluding medical schools. being 0.64 and 0.70, respectively. Coelli[16] assessed the efficiency of Australian HEIs using cross-sectional data for 1994. The findings indicated that the mean technical efficiency scores for the Australian university model were quite high, 0.95. In addition, Athanassopoulos and Shale[9] applied DEA to estimate the efficiencies of 45 established universities in the UK during 1992/93. Their findings showed that, in the cost efficiency model, the mean efficiency for all institutions was estimated to be 0.83. Later, in 1992/93, McMillan and Datta[28] employed DEA to evaluate the efficiency of 45 Canadian universities. The results from these DEA analyses show that the mean efficiency score for universities with medical schools was 0.94, whereas that for universities without medical schools was 0.95.

For the period 2000–2015, studies on the efficiency of HEIs have continued to develop in the advanced countries with deeper analyses and have expanded in European and Asian nations. For example, Avkiran[10] applied a DEA method to measure the efficiency of 36 Australian universities based on a 1995 dataset. His findings were that the mean efficiency score was 0.96 for the overall model, 0.97 for the model of delivery of services, and 0.63 for the fee-paying enrolments model. A third study published by Abbott and Doucouliagos[1] estimated the performance of universities, using the same 1995 data as used in the work of Avkiran.[10] Their findings showed that, the average technical and scale efficiency scores were 0.95 and 0.97, respectively. Carrington, Coelli, and Rao[13] measured productivity growth by using DEA methods for 35 Australian universities with annual data over the period 1996–2000. The results suggested that universities were relatively efficient and that their efficiency was stable over the period.

By investigating the possibility of measuring the efficiency of HEIs, Johnes[26] asserted that, with the ability to treat many inputs and outputs, DEA was a striking technique for measuring the performance of HEIs. She added, however, its disadvantages should be concerned. The author applied DEA to a dataset of more than 100 British HEIs using data for the academic year 2000-2001. The findings indicated that technical efficiency and scale efficiency in the British HEIs appeared to be high, on average at 0.946. The bootstrapping measures suggested that there was a significant difference in efficiency of the worst- and best-efficient HEIs.

More recently, many studies have tended to analyse the performance of HEIs in different nations at the institutional and national levels. Agasisti and Pohl[6] used a two-stage analysis to examine and compare the efficiencies of 53 Italian and 69 German public universities and their evolution for the years 2001–2007, respectively. Their findings from the CRS DEA model indicated that the mean efficiency score within the Italian universities was only 0.69, whereas, for the German dataset, the average was 0.77.

One of the first papers using DEA in research on the efficiency of HEIs in Asian countries was that of Castano and Cabanda.[14] They estimated the efficiency and productivity growth of 59 HEIs (State Universities and Colleges, SUCs) in the Philippines over the period 1999–2003. Findings using the Malmquist index model revealed that 49 SUCs were efficient whereas six SUCs showed technological progress. The mean technical efficiencies using the CRS- and VRS-DEA models and scale efficiency were 0.95, 0.97, and 0.99, respectively. Another study by Johnes and Yu[25] used DEA to examine the relative efficiency of over 100 selected universities from the People's Republic of China using data for 2003 and 2004. Their findings indicated that the level of efficiency depended on the presence of a subjective measure of research output in the model. When the reputation variable (based on experts' opinions) was included, the mean efficiency was higher at 0.90, but when it was excluded it was approximately 0.55.

The recent work of Husain[24] who investigated the efficiencies of 20 Malaysian public universities for the period 2006–2008 revealed that the average university efficiency score was 0.87 using the VRS-DEA, and 0.74 using the hybrid returns-to-scale (HRS) with trade-offs method. The author asserted that the latter increased the discriminatory power of the DEA assessment as reflected by the lesser number of universities identified as efficient, and smaller efficiency scores. Later, using the same dataset of 20 public universities in Malaysia, Podinovski

DOI: https://doi.org/10.30564/jiep.v1i1.380

and Husain[32] emphasised the importance of the new method in which the HRS DEA model is integrated with the usage of production trade-offs in higher education. However, their method requires the critical assumption of selective proportionality (between student and staff in their study). Furthermore, Duh et al.[21] investigated the operational efficiencies of 99 Taiwanese universities using cross-sectional data for 2005. Their results showed the average efficiencies of 0.78 and 0.48 for teaching and research operations, respectively.

Nguyen, Thenet and Nguyen[30] applied the traditional DEA method to estimate 30 Vietnamese universities' performance, using a single dataset for the academic year 2012/13. The authors did a sensitive analysis by testing eight various models in terms of different inputs and outputs. The results indicated that the scores of efficiencies of these universities ranged from 0.81 to 0.92. However, this study used quite a small sample size for a cross-sectional data; thus, this would lead to lack of analysis power, causing less robust findings.

Tran and Villano[42] investigate the performance of 50 universities and 50 colleges in Vietnam using the two-stage DEA model. Their findings showed that the average efficiency score of colleges was quite high at 0.94 and external factors affected the inefficiency of colleges. However, because the sample size of this paper is not large enough, the efficiency scores might be overstated. In addition, this paper only used a cross-sectional data that could not capture change of colleges' efficiency over time.

More recently, Tran and Villano[43] applied the advanced dynamic network DEA to estimate efficiency of 116 Vietnamese state colleges for 2011/12–2013/14. They indicated that these colleges were not efficient in their operations and that colleges were efficient only if they were efficient at individual financial and academic divisions, respectively, the mean efficiency being 0.74. This paper used a panel data to capture change in efficiency of colleges over time but did not test the effects of environmental factors against inefficiency of colleges, therefore left this as an open question. Our paper comes to fill this gap by examining the performance of colleges over multiple periods and determining determinants of inefficiency of Vietnamese colleges.

By doing this, our research on determinants of colleges' academic performance would bring benefits by: (a) providing insightful information about operational efficiency of colleges varying over time; (b) identifying determinants of colleges' inefficiency; and (c) recommending appropriate policies to improve colleges' performance towards innovation.

4. Methodology

4.1 Data Envelopment Analysis (DEA)

DEA is a linear programming approach for estimating the performance of decision-making units, which allows to convert multiple inputs into multiple outputs.[17,25,42] DEA is widely applied in different sectors, especially in higher education where the price is not available.[25] Thus, this approach is preferred to estimate the efficiencies of tertiary education institutions. Nevertheless, the shortage of this method is that it does not allow one to distinguish managerial inefficiency from random noise. [42]

DEA is an appropriate method applied in context of Vietnamese colleges. This is because (1) several factors are not estimated in monetary terms such as quality of students or qualifications of staff;[10,13,21] (2) many policies recently issued affect colleges in different ways and thus make it difficult to identify clear direction of influences; and (3) DEA can differentiate inefficient and efficient colleges, therefore would help erect performance targets for inefficient DMUs.

DEA was officially coined by Charnes, Cooper and Rhodes[15] to estimate technical efficiency of a DMU. This method uses multiple inputs and multiple outputs and assumes constant return to scale (CRS) in which all DMUs operate at an optimal scale. Nevertheless, it is widely recognised that the optimal operating scale of DMUs can be affected by some external factors, for example, imperfectly competitive status, government rules, financial challenges, etc.[17] Thus, the CRS DEA method could not capture these changes. Later, Banker, Charnes and Cooper[12] proposed an extension to this model to demonstrate the variable returns to scale (VRS) situation for the purpose of estimating technical efficiency without scale effects.

This model is applied to Vietnamese colleges and depicted as below:

For each college, $X = (x_1, \ldots, x_m)$ is a vector discretionary inputs used. $Y = (y_1, \ldots, y_s)$ is a vector of outputs produced. Accordingly, $X_j = (x_{1j}, \ldots, x_{Mj})$ and $Y_j = (y_{1j}, \ldots, y_{sj})$ are inputs and outputs for college j. The production possibility set is illustrated by

$$L_v = \left\{ (X,Y);\ \sum_{i=1}^{N} \lambda_i y_{si} \geq y_s, \sum_{i=1}^{N} \lambda_i x_{mi} \leq x_m, \sum_{i=1}^{N} \lambda_i = 1, \lambda_i \geq 0 \right\} \quad s = 1, \ldots, S; m = 1, \ldots, M; i = 1, \ldots, N \tag{1}$$

where λ_i are coefficients to increase or decrease the observed operations of college i ($i = 1,\ldots,N$) with respect to inputs (x_i) and outputs (y_i). θ is the efficiency scores that can be computed by the following linear programming problems:

 DOI: https://doi.org/10.30564/jiep.v1i1.380

$$\hat{\theta}_{VRS} = min\{\theta > 0 | y_s \leq \sum_{i=1}^{n} \lambda_i y_{si}, \theta x_m \geq \sum_{i=1}^{n} \lambda_i x_{mi}, \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0, i = 1, \dots n\} \quad (2)$$

where $\hat{\theta}_{VRS}$ is the prediction of an observed college (x,y) to the efficient frontier, and give the initial technical efficiency of the i^{th} college. For all $(x,y) \in L_v, \hat{\theta}_{VRS} = 1$, the college is fully technical efficient if $\hat{\theta}_{VRS} = 1$.

Sampling variation and random errors are not taken into account in (2), thus it could cause potential biases in the results of efficiency scores.[37] This can be solved by using the bootstrap technique proposed by Simar and Wilson.[36] This technique is used in this paper to generate more robust results for efficiency scores.

The DEA input orientation is opted to estimate the efficiency of colleges. This choice is suitable with the Vietnamese context where colleges operate within the confines of regulations. This means that they could not expand their output without meeting strict regulations from the government. Instead, they should use their available input resources efficiently to obtain the existing registered output.

It is generally recognized that the DEA results can be sensitive to outliers because this method is a deterministic frontier approach without accounting for random disturbances. Therefore, we first attempted to determine the outliers, then proceeded with conducting efficiency analysis of colleges. The process has been conducted following the methods of Andersen and Petersen,[8] Thanassoulis, [40] Thanassoulis et al.[41] The details of this process have not been presented here but is available in Tran and Villano.[43] As a result, nine (9) college outliers were identified and are kept sitting on the boundary drawn on non-outlier colleges so that they would not affect the position of the efficiency boundary.

4.2 Determinants of Colleges' Performance

Examining the effects of environmental factors on DEA efficiency in the second stage has received much attention of researchers regarding econometric models used. It is widely recognised that the ordinary least squares (OLS) model is inappropriate because the prediction of the dependent variable may exceed the range of zero and one and its estimated coefficients do not reflect the nature of DEA bounded scores and the presence of many points at one in their distribution. On the other hand, the Tobit model used for the second-stage DEA analysis is also questionable. It is observed that observations at one is a result obtained from the way DEA scores are defined, not result of the Tobit censoring. Additionally, the DEA efficiency scores of zero are not observed as well. In other words, the domain of the two-limit is not similar to that of the DEA scores.[33,38]

Using the logit factional regression model proposed by Papke and Wooldridge[31] the second-step DEA method has been implemented recently in studies of Hoff [23] and McDonald.[27] By comparing various approaches for modelling the second stage of DEA, these authors supported the use of the simple regression analysis. However, McDonald[27] acknowledged the advantages of the Papke and Wooldridge's model to obtain more robust estimates.

The study of Simar and Wilson[38] was the first to describe a coherent data-generating process (DGP) for DEA scores. They provided a set of assumptions in which the use of estimates (not true efficiency scores) does not affect the consistency of the second stage regression parameters. Simar and Wilson[38] proposed two alterative bootstrap methods, Algorithm 1 (without taking account of the bias term in the first stage) and Algorithm 2, accounting for the sampling variability of DEA scores to make a valid statistical inference about these parameters.

Banker and Natarajan[12] later proposed a straight-line relationship between the log of efficiency scores and the environmental variables. They imply that using the linear regression analysis can generate consistent parameters in the second stage of DEA method. However, as compared to DGP of Simar and Wilson,[38] their methods is quite constrained. Additionally, since the predicted variable is the log, the its value needs to be re-estimated, thus it is not actually the level of DEA scores per se.

More recently, Ramalho, Ramalho and Henriques[33] proposed several alterative regression models of efficiency scores in the second stage using fractional regression models and tests of the specification chosen for the regression model using simple statistical tests. They also suggested that two-part fractional regression models may be beneficial when the propotion of unity values is large. However, the method of Ramalho, Ramalho and Henriques[33] did not consider the sampling variability of DEA scores as in the method of Simar and Wilson,[38] thus did not solve the problem of inferences about the regression parameters. In this sense, the method of Simar and Wilson[38] appears to the only feasible way to provide valid inference for regression parameters. [33]

All things considered, the method of Simar and Wilson[38] seems to be most appropriate to yield a valid inference for the parameters of interest in the second stage regression model. In this study, we adopt the Simar and Wilson[38]'s model, Algorithm 2 for the second stage DEA analysis to examine environmental impacts on efficiencies of colleges. The details of this process have not been shown here but are available in Tran and Villano.[42]

DOI: https://doi.org/10.30564/jiep.v1i1.380

5. Data and Variables

5.1 Output and Input Variables

The paper seeks to offer the best way to quantify the outputs of education. The first output is the number of full-time equivalent (FTE) students. This variable has been utilised in previous research including Abbott and Doucouliagos,[1,2] Thanassoulis, et al.,[41] de Miranda, Gramani and de Carvalho Andrade,[20] and Daghbashyan.[18] On the other hand, some articles have used this output as an input to produce the graduate outputs, e.g. Agasisti and Johnes,[5] Agasisti and Bianco,[4],Agasisti and Pohl.[6] However, it can be seen that the outputs in higher education are more special than in other fields because we would train students for a long term of 3 or 4 years until they can get their degrees. This means that input resources such as staff, facilities and operating costs would be used to produce the student outputs for the whole study phase of students. Using students as inputs to produce the graduate outputs may be only relevant in the case that students who are studying in the final year will graduate in that year, given that their previous learning years should be ignored. However, to our knowledge, it would be better to see the annual enrolments as outputs rather than inputs because this would reflect the real nature of higher education.

The second output is the number of completed students per annum. This is used as a proxy to estimate the quality of teaching. Previous studies such as Stevens[39] and Daghbashyan[18] used the ratio of first- and second- class degree, or employment possibilities after graduation as a proxy for teaching quality. However, in the case of Vietnam, quality of teaching output has not been transparent and hard to be evaluated in an exact way. It is noted that the number of graduates per annum should be separated from the number of students (enrolments) who are still studying at that time. This would ensure that these two kinds of outputs reflect the adequateness of the model specification.

The final output refers to incomes obtained from research-based activities (not consisting of tuition fees and government funding). Previous research, e.g. Robst,[34] Abbot and Doucouliagos,[1] Daghbashyan[18] used research funding as a surrogate variable for research output though their choice did not take into account for quality and disciplinary differences. For the case of Vietnamese colleges, because of limited data, the incomes from research-related operations are used as a proxy for research output.

Similar to the work of Tran and Villano,[43] four input variables are used in our paper: (1) total number of academic staff (full-time equivalent, FTE) who take charge of teaching and research activities, (2) the number of non-academic staff (FTE) who are involved with supporting activities for learning and teaching and management at different levels; (3) floor area for academic spaces, a basis of calculating annual enrolment quotas; and (4) operating cost, annual expenditure for academic operations.

5.2 Effects of External Factors on Efficiency of Colleges

We choose to examine the external variables that may influence the performance of colleges in the second-step DEA method using the truncated regression analysis including age of the college, place (in main cities or outside main cities), the share of academic staff with postgraduate and undergraduate degrees, tuition revenue proxied as a key financial resource of colleges. The following hypotheses are tested in the DEA second-step regression analysis as follows.

Place is projected to have a positive relationship with efficiencies of colleges. Metropolitan colleges have more plus points since they can access input resources more easily for academic operations.

Type or ownership of colleges is predicted to be positively related with the efficiencies of colleges. Public colleges may be more advantageous as they receive more funding from the government.

The average national entry exam marks (NEE) are anticipated to influence the efficiencies of HEIs. Its effects can be positive or negative as the higher NEE can make a decrease in both the number of enrolments of colleges and staff.

Age or establishment of colleges is projected to has a positive association with the efficiency scores because the older colleges have a good reputation and their leaders have much more experience in managing schools; therefore, their performance may be better.

The proportions of postgraduate and undergraduate staff, respectively, are expected to be positively related with colleges' performance. Note that these effects are relative to those staff with no formal academic degrees. These variables are tested separately to avoid the serial correlation.

Tuition fees are forecast to have a positive and significant relationship with the efficiency scores, especially the years following Decree 49/2010/ND-CP of the government issued in 2010.

5.3 Data Sources

A total of 141 colleges for 2011/12-2013/14 are used in this study with data sources from MOET where performance indicators of colleges have been recorded in terms of their annual reports. These colleges complied with the rules of MOET to send in their annual reports for all three years. Our surveyed sample accounts for 53 % of the total number of Vietnamese colleges. While a long span of data

 DOI: https://doi.org/10.30564/jiep.v1i1.380

is desirable, a panel data of the three-year period in this paper fundamentally meets requirements to capture a variation in colleges' efficiency over multiple periods.

A summary on inputs, outputs and control variables is presented in Table 2. On average, colleges trained 2,113 students and the number of graduated students was 554. The income for research-related activities is averagely 3.61 billion VND. The non-academic staff is less than the number of academic staff, 54 and 129 respectively. Depending on the training size of colleges, there is a variation in using input resources to produce outputs. In addition, the annual enrolment quotas approved by the government are based on the ratio of students to academic staff and the ratio of students to floor area for academic spaces.

6. Empirical Findings

This section illustrates the estimated findings of the efficiencies of colleges for the period 2011/12–2013/14 and investigates determinants of colleges' performance.

6.1 Academic Efficiencies of Colleges

This section presents the results of colleges' efficiencies varying over three years, using the conventional and bootstrapped approaches. As can be seen in Table 3, the average efficiency of the colleges in the sample is 0.697 and 0.623 for in the conventional and bootstrapped approaches, respectively. In other words, colleges need to enhance their performance by 30.3% and 37.7% in terms of these models. However, this result is much lower than the result of Tran and Villano[42] at 0.939. This can be explained by a difference in the sample size and a type of data used. While Tran and Villano[42] used a cross-section data and a relatively small sample size of 50 colleges, we have used a panel data for three years with a larger sample size of 141 colleges that would technically produce more robust findings. Overall, our result lies within the range of colleges' efficiencies scores as summarised in Tran.[44] 10 out of total colleges are fully efficient in the standard model, and the remaining colleges are operating below the frontier efficiency, thus more improvement is necessary for their academic operations. However, no colleges are found to be fully technically efficient in the bootstrapped model. This result is in line with Tran and Villano.[42]

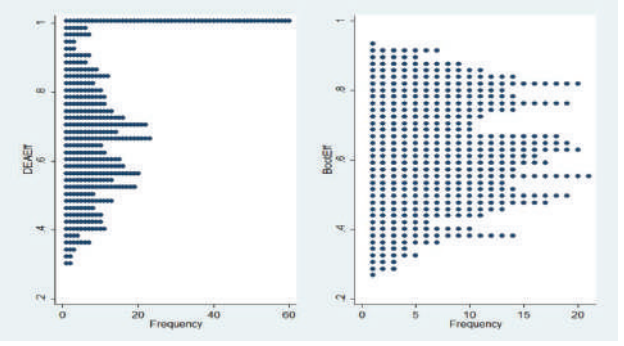

Figure 1. Frequency Distribution of Efficiencies of Colleges

Table 2. Statistics Summary on Input and Output variables

	Unit	Mean	SD	Minimum	Maximum
Outputs					
Undergraduates	Student	2112.93	1498.62	83	7116
Completed students	Student	554.37	398.84	14	1623
Research income	Billion VND	3.61	7.7	0.001	55
Inputs					
Floor area [a]	1000 m^2	13.29	0.81	0.12	4.22
Non-academic staff	Person	54.31	20.78	25	160
Academic staff	Person	129.28	67.11	32	438
Operating costs	Billion VND	18.38	12.85	0.9	67.76
External factors [b]					
Age	Year	12.09	6.52	4	39
NEE	Mark	11.47	2.08	10	21.75
Ratio of Postgraduate staff	Percentage	0.428	0.143	0.063	0.779
Ratio of Undergraduate staff	Percentage	0.572	0.143	0.221	0.937
Tuition revenue	Billion VND	8.39	9.31	0.038	70.22

Notes: [a] Floor area for academic spaces (classroom, library, etc…); [b] Excluding location and type that are dummy variables.

DOI: https://doi.org/10.30564/jiep.v1i1.380

Table 3. Efficiencies of Colleges over Three Years

	Standard VRS Efficiency				Bootstrapped VRS Efficiency			
	2011	2012	2013	Overall	2011	2012	2013	Overall
Mean	0.717	0.680	0.694	0.697	0.639	0.610	0.619	0.623
SD	0.197	0.199	0.184	0.175	0.163	0.167	0.146	0.143
Min	0.295	0.305	0.337	0.325	0.258	0.277	0.308	0.297
Max	1	1	1	1	0.925	0.916	0.917	0.891
Eff.units [a]	24	17	18	10	0	0	0	0
Hotelling test [b] (F value)				1051***				

Notes: [a] The number of colleges with efficiency scores of 1; *** denotes significance at 1%; [b] Hotelling test for equal means standard and bootstrapped DEA scores.

Figure 1 illustrates the histogram of efficiencies of colleges over three years. The frequency of efficiencies of colleges in the standard DEA model is quite dense and focuses on the value of one, whereas that in the bootstrap model is sparse and less than one. This implies that after isolating noises, the efficiencies of colleges decrease and reflect nearly the nature of their performance.

It is observed from Table 4 that public colleges are less efficient than private colleges. The average score of public colleges is 0.614 whereas that of private ones is 0.692 in the bootstrapped model. This distinction is significant at the 1% significance level. This finding is accordant with Tran and Villano.[42] It may reflect the fact that investment capitals in private colleges have been used in a more efficiency way for academic activities. Interestingly, the finding in Tran and Villano[43] showed that the efficiency of public college is 0.74, being slightly higher than our finding.

Figure 2 illustrates a moving trend to the right near one with the higher efficiencies for private colleges each year and over years after implementing a bootstrap procedure. By contrast, the efficiencies of public colleges have a downward trend to the right with the lower efficiencies.

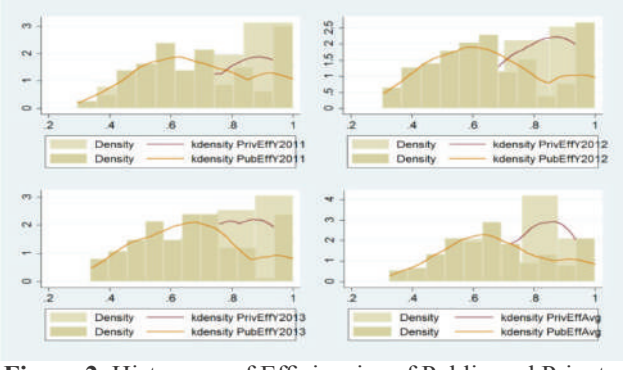

Figure 2. Histogram of Efficiencies of Public and Private Colleges

6.2 Determinants of Colleges' Inefficiencies

The influences of environmental factors on the performance of colleges are investigated in the second-step DEA approach. It is noted that since the Shephard scores obtained the first step are greater than one indicating the inefficiency level of colleges. The sign of coefficients in the second-step regression should be interpreted inversely on efficiencies of colleges, meaning that positive signs affect negatively the performance of colleges and vice versa. We tested two models, Model 1 and Model 2, for

Table 4. Efficiencies of Colleges Classified by Ownership

	Standard VRS Efficiency				Bootstrapped VRS Efficiency			
	2011	2012	2013	Overall	2011	2012	2013	Overall
Public	0.705	0.670	0.680	0.685	0.633	0.602	0.607	0.614
Private	0.810	0.764	0.803	0.792	0.690	0.677	0.710	0.692
Wilcoxon rank-sum test [a]				3.71***				3.328***

Notes: [a] The Wilcoxon rank-sum test for equal means between scores of public and private HEIs.

 DOI: https://doi.org/10.30564/jiep.v1i1.380

individual cases of colleges. The main difference between these models is independent variables of the proportions of academic staff with postgraduate and undergraduate degrees. These two variables are examined independently to avoid the serial correlation phenomenon in regression models. The results of these two models are presented in Table 5.

In both models, most variables are significantly different from zero at 1% level of significance. Place with a positive sign implies that metropolitan colleges are more efficient than their rural counterparts because the former can be more beneficial in accessing input resources and, thus, probably have more enrolments. This result is associated with the findings of Carrington et al.[13] and Tran and Villano.[42] It is interesting that NEE is positively correlated to the efficiencies of colleges. This could be explained that the higher the NEE, new enrolments can be declined to some extent, thus colleges may use fewer of their input resources relative to their outputs, e.g. academic staff. As a result, their efficiencies can be improved. Unexpectedly, both age and tuition fees are not correlated to the efficiencies of colleges. It can be observed that in the years following Decree 49/2010/ND-CP, revenues on tuition fees of colleges have not been a key enhancer to enhance the efficiencies of colleges.

The findings in Table 5 show that the ratio of undergraduate staff is positively related to the efficiencies of colleges, whereas that of postgraduate staff has a negative coefficient. Albeit these influences are not significant, they imply the fact that academic staff with a bachelor degree seem to be enough for colleges. This is because at the college level, the importance of teaching activities has been emphasised rather than that of research activities. In addition, a higher proportion of postgraduate staff (Master or PhD degree) would increase expenditure for colleges. Thus, academic staff with undergraduate degree still occupies a vital role in colleges.

7. Discussion and Conclusion

Our paper made efforts to conduct a performance measurement of Vietnamese colleges with panel data for 141 colleges during the period of 2011/12–2013/14. The second stage bootstrapping DEA approach is suggested to investigate the performance of colleges and the influence of external factors on their performance. The empirical results are anticipated to provide more information for policy makers and educational managers to design more appropriate policies to move performance of colleges forward.

Using the second stage bootstrapping DEA model, the findings indicate that the efficiency of colleges in the surveyed sample are not high as expected, at 0.697 and 0.623 for the conventional and bootstrapped models. The potential improvement is necessary for colleges to obtain the full efficiency of unity by using input resources more appropriately. Unexpectedly, public colleges are less efficient than their private counterparts, 0.614 and 0.692, respectively. The location and NEE are external factors affecting positively the efficiencies of colleges. By contrast, revenue from tuition fees and the ratio of postgraduate

Table 5. Influences of environmental factors on the performance of colleges

	Model 1 (with the ratio of postgraduate staff)				Model 2 (with the ratio of undergraduate staff)			
	Coefficient	Bootstrap std.error	z	P>|z|	Coefficient	Bootstrap std.error	Z	P>|z|
Location	-0.403***	0.133	-3.040	0.002	-0.408***	0.13	-3.12	0.002
Type	0.520***	0.192	2.710	0.007	0.52***	0.19	2.67	0.008
NEE	-0.061**	0.025	-2.380	0.017	-0.061**	0.025	-2.44	0.015
Age	0.0066	0.0116	0.570	0.571	0.0066	0.0115	0.57	0.566
Ratio Poststaff	0.614	0.384	1.600	0.110				
Ratio Understaff					-0.614	0.387	-1.59	0.113
Tuition	0.023	0.015	1.560	0.120	0.023	0.015	1.52	0.129
Age*Tuition	-0.0015	0.0011	-1.360	0.174	-0.0015	0.0011	-1.37	0.169
Constant	1.343***	0.340	3.950	0.000	1.96***	0.43	4.53	0
σ^2	0.697***	0.053			0.697***	0.0527		
Wald χ^2	24.16				25.18			
p value > χ^2	0.001				0.0007			

Note: **, *** denotes significance at the 5% and 1% levels; 2000 replications for bootstrapping

DOI: https://doi.org/10.30564/jiep.v1i1.380

staff were not significantly associated with the efficiencies of colleges.

The above findings lead us to some managerially practical implications. In the first place, given their educational quality, the surveyed colleges were operating below the frontier efficiency. The efficiencies of colleges witnessed a volatility over the three years involved. In fact, from 2011, MOET suggested universities and colleges reduce the enrolment quotas in the business discipline. This is since the number of unemployed graduates of this discipline were high and led to an imbalance in the labour market. Consequently, the enrolment quotas for these study fields have gone down in the following years. This has led to influence financial resources of colleges via tuition fees, especially public colleges in those years, even though public colleges are allowed to increase the level of fees greater than that of previous years. The decreased enrolments and financial resources might cause a decrease in the efficiencies of colleges. However, drawing a conclusion of whether this decline comes from the result of policy intervention is out of this paper due to the complexity of identifying deterministic and causal associations in this context.

Second, albeit the shares of postgraduate and undergraduate staff did not illustrate their significant impacts on the efficiencies of colleges, increasing the share of postgraduate academic staff, especially PhD qualification should be concerned to enhance academic operations of colleges. This would help colleges accelerate innovation in learning and teaching, thus enhance education quality. Accordingly, support from the government is necessary to provide them, especially to public colleges, with enough flexibility in managerial mechanism of physical and human capital resources to facilitate the process of innovation in teaching and learning.

Finally, interestingly, tuition revenues did not contribute to the efficiencies of colleges. Although the government policy has facilitated colleges to increase their revenue via increasing tuition fees within a given framework of 2010/11-2014/15, the growth in their revenues was quite marginal and not enough to contribute to their performance. In this sense, a further examination of this policy may be useful to help policy makers have a better policy in the future.

This paper has bridged the gap in the literature by investigating changes in the performance of colleges over time and examining determinants of colleges' inefficiency for the period of 2011/12–2013/14. Some following aspects might be implemented to make our findings more comprehensive. First, our sample is 53 % of total numbers of colleges in Vietnam for a three-year period, thus more observations with more periods of data would be necessary to provide more insights of volatility in productivity of colleges over longer period. Together with this, more inputs and outputs can be added in further studies, for instance, publications of academic staff. In addition, the quality of graduates, such as their study record or evaluation of employers, should be also taken into account. Further, qualifications of academic staff should be measured by actual numbers and weighted by different levels. Finally, in the Vietnamese context, environmental factors could impact directly on input usages and thus cause a decrease in the efficiencies of colleges. Future studies should account for this point to supplement our findings.

References

[1] Abbot M, Doucouliagos C. The efficiency of Australian universities: A data envelopment analysis. Economics of Education Review, 2003, 22:89-97.

[2] Abbot M, Doucouliagos C. Competition and efficiency: overseas students and technical efficiency in Australian and New Zealand universities. Education Economics, 2009, 17:31-57.

[3] Agasisti T, Catalano G, Landoni P, Verganti R. Evaluating the performance of academic departments: An analysis of research-related output efficiency. Research Evaluation, 2012, 21:2-14.

[4] Agasisti T, Bianco AD. Reforming the university sector: Effects on teaching efficiency—— Evidence from Italy. Higher Education, 2009, 57:477-498.

[5] Agasisti T, Johnes G. Beyond frontiers: Comparing the efficiency of higher education decision-making units across more than one country. Education Economics, 2009, 17:59-79.

[6] Agasisti T, Pohl C. Comparing German and Italian public universities: Convergence or divergence in higher education landscape? Managerial and Decision Economics, 2012, 33:71-85.

[7] Ahn T, Charnes A, Cooper WW. Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. Socio-Economic Planning Sciences, 1988, 22(6):259-269.

[8] Andersen P, Petersen NC. A procedure for ranking efficient units in Data Envelopment Analysis. Management Science, 1993, 39: 1261-1264.

[9] Athanassopoulos AD, Shale E. Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis. Education Economics, 1997, 5(2): 17-134.

[10] Avkiran NK. Investigating technical and scale efficiencies of Australian Universities through data envelopment analysis. Socio-Economic Planning Sciences, 2001, 35:57-80.

 DOI: https://doi.org/10.30564/jiep.v1i1.380

[11] Banker RD, Charnes A, Cooper WW. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science, 1984, 30: 1078-1092.

[12] Banker RD, Natarajan R. Evaluating contextual variables affecting productivity using data envelopment analysis. Operations Research 2008, 56:48-58.

[13] Carrington R, Coelli T, Rao P. The performance of Australian Universities: conceptual issues and preliminary results. Economic Papers, 2005, 24:145-163.

[14] Castano MCN, Cabanda E. Sources of efficiency and productivity growth in the Philippine state universities and colleges: A non-parametric approach. International Business and Economics Research Journal, 2007, 6:79-90.

[15] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. European Journal of Operational Research, 1978, 2:429-444.

[16] Coelli TJ. Assessing the performance of Australian universities using data envelopment analysis. Mimeo 1996, Centre for Efficiency and Productivity Analysis, University of New England.

[17] Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE. An Introduction to Efficiency and Productivity Analysis 2005, 2nd Edition. New York: Springer.

[18] Daghbashyna Z. The economic efficiency of Swedish higher education institutions. Working paper 245 2011, the Royal Institute of Technology, CESIS.

[19] Dan Tri. Năm 2017: Chi ngân sách cho giáo dục là 248.118 tỷ đồng (2017: Expenditures for Education and Training are 248,118 VND) 2018;
Accessed 26 November 2018 from: https://dantri.com.vn/giao-duc-khuyen-hoc/chi-ngan-sach-cho-giao-duc-la-248118-ty-dong-20180930163940791.htm.

[20] De Miranda RO, Gramani MCN, de Carvalho Andrade E. Technical efficiency of business administration courses—— A simultaneous analysis using DEA and SFA. Insper Working Paper WPE: 214, 2010. Rua Quata, Sao Paulo-SP, Brazil: Institute of Education and Research.

[21] Duh RR, Chen KT, Lin RC, Kuo LC. Do internal controls improve operating efficiency of universities? Annals of Operations Research, 2014, 221:173-195.
21. Emrouznejad A, Thanassoulis E. A mathematical model for dynamic efficiency using data envelopment analysis. Applied Mathematics and Computation, 2005, 160:363-378.

[22] Fried HO, Lovell CAK, Schmidt SS. Efficiency and productivity. In Fried, H.O., Lovell, C.A.K., and Schmidt, S.S. (Eds.): The Measurement of Productive Efficiency and Productivity Growth, 2008, 3–91. New York, USA: Oxford University Press, Inc.

[23] Hoff A. Second stage DEA: Comparison of approaches for modelling the DEA scores. European Journal of Operational Research, 2007, 181:425-435.

[24] Husain WRW. Comparative Evaluation of Public Universities in Malaysia using Data Envelopment Analysis 2012. PhD Thesis, Warwick Business School. The University of Warwick, Conventrance, England.

[25] Johnes J, Yu L. Measuring the research performance of Chinese higher education institutions using data envelopment analysis. China Economic Review, 2008, 19: 679-696.

[26] Johnes J. Measuring teaching efficiency in higher education: An application to data development analysis to economic graduates from UK universities, 1993. European Journal of Operational Research, 2006, 174:443-456.

[27] McDonald J. Using least squares and Tobit in second stage DEA efficiency analyses. European Journal of Operational Research, 2009, 197:792-798.

[28] McMillan ML, Chan WH. University efficiency: A comparison and consolidation of results from stochastic and non-stochastic methods. Education Economics, 2006, 14(1):1-30.

[29] MOET (Ministry of Education and Training). Statistics on education; Accessed on 27 November 2018 from: https://moet.gov.vn/thong-ke/Pages/thong-ko-giao-duc-dai-hoc.aspx.

[30] Nguyen TTH, Thenet G, Nguyen KM. Applying DEA sensitivity analysis to efficiency measurement of Vietnamese universities. Management Science Letters, 2015, 5:983-992.

[31] Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401 (K) plan participation rates. Journal of Applied Econometrics, 1996, 11:619-632.

[32] Podinovski VV, Husain WRW. The hybrid returns-to-scale model and its extension by production trade-offs: An application to the efficiency assessment of public universities in Malaysia. Annals of Operations Research, 2015, 250(1):65-84.

[33] Ramalho EA, Ramalho JJS, Henriques PD. Fractional regression models for second stage DEA efficiency analyses. Journal of Productivity Analysis, 2010, 34:239-255.

[34] Robst J. Cost efficiency in public higher education institutions. Journal of Higher Education, 2001, 72:730-750.

[35] Sav GT. Managing operating efficiencies of publicly owned universities: American university stochastic frontier estimates using panel data. Advances in Management and Applied Economics, 2012, 2:1-23.

[36] Simar L, Wilson PW. Sensitivity analysis of efficiency score: How to bootstrap in non-parametric frontier models. Management Science, 1998, 44:49-61.

[37] Simar L,Wilson PW. Bootstrapping in non-parametric

  DOI: https://doi.org/10.30564/jiep.v1i1.380

frontier models. Journal of Applied Statistics, 2000, 27:779-802.

[38] Simar L, Wilson PW. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 2007, 136:31-64.

[39] Stevens PA. A stochastic frontier analysis of English and Welsh universities. Education Economics, 2005, 13:355-374.

[40] Thanassoulis E. Setting achievement targets for school children. Education Economics 1999, 7:101-119.

[41] Thanassoulis E, Kortelainen M, Johnes G, Johnes J. Costs and efficiency of higher education institutions in England: A DEA analysis. Journal of the Operational Research Society, 2011, 62: 1282-1297.

[42] Tran CDTT, Villano RA. An empirical analysis of the academic performance: The case of Vietnamese higher education institutions. Journal of Further and Higher Education, 2017, 41(4):530-544

[43] Tran CDTT, Villano RA. Measuring efficiency of Vietnamese public colleges: An application of the DEA-based dynamic network approach.International Transactions in Operational Research, 2018, 25(2):683-703.

[44] Tran CDTT. The reform process of higher education in Vietnam: Performance efficiency and transparent accountability. Review of Educational Theory (2018), 1(4):121-133.

[45] Vietnamnet. 5 năm, ngân sách cho giáo dục tăng 92.500 tỷ đồng (For five years, state budget for education increases 92,500 VND) 2018;
Accessed 26 Nov 2018, Available from: http://vietnamnet.vn/vn/giao-duc/nguoi-thay/5-nam-ngan-sach-cho-giao-duc-tang-92-500-ty-dong-482719.html.

 DOI: https://doi.org/10.30564/jiep.v1i1.380

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep/

ARTICLE

Holonic Theory and Holistic Education

Josep Gallifa*
FPCEE Blanquerna. Ramon Llull University, C. Cister 34, 08022 Barcelona, Spain

ARTICLE INFO

Article history:
Received: 18 December 2018
Accepted: 27 December 2018
Published: 18 January 2019

Keywords:
Holonic Theory
Holistic Education
Holistic Pedagogy
Phenomenology

ABSTRACT

This paper presents the holonic theory, which is an attempt to develop in a single model the explanation of the evolution in the physic, biologic and cultural dimensions. The purpose of this development is to understand the traits of nowadays common holism, which is considered a necessary thinking practice in different domains, among them education. A phenomenological study has been developed connecting diverse noospheric holons. The results allow for a characterization of the holonic structure of a holistic consciousness act. This characterization is used to define holistic education: An education with the dimensions of preservation, profundity, projective action and span. This article can also be interpreted as a contribution to provide the ontological and epistemological bases for going beyond the modern and postmodern worldviews. These findings help in providing a framework for today's holistic pedagogical debates and developments.

1. Introduction

In the most part of professions that involve human relationships and, more generally, in contemporary intellectual perspectives there is a common need: to think comprehensively, globally, holistically. Effectively, today's problems rarely can be approached only from only one specialty, and integrated or integral approaches are needed in order to grasp the complexity of issues and phenomena. Education is in the midst of this contemporary trend. Nevertheless to create holistic or integral kinds of thinking there is a previous need to be addressed and focused on, which is to have a well-defined ontological model, able to accommodate integrated and holistic -instead of fragmented- perspectives. Therefore a first question to be discussed is: Is there an ontological system as inclusive as possible of multiple perspectives?

On the other hand, there is another important need in intellectual and practical endeavors that can be formulated, in a simple way, as recovering the 'subject', by presenting a suitable model of consciousness. The bias towards the 'objective' knowledge let the legacy of premodern traditions aside. Recovering the 'subject' in the context of the nowadays modern and postmodern worldviews is thus another related important need. Wilber[1] presented an integral theory to integrate these worldviews and traditions in a single model. Particularly meaningful is the contribution of Wilber presenting a model for the evolution of the subject, in the form of a spectrum of consciousness, in which he described the levels of evolution of the subjective consciousness. The model of Wilber was used to substantiate in turn different approaches: Beck and Cowan in the value-memes of a culture,[2] Esbjörn-Hargens or Murray in integral education,[3] Laloux in the evolution of organizations,[4] Thomas in integral leadership,[5] among many others. All of these approaches are founded on subjective acts of holistic consciousness. Another question,

**Corresponding Author:*
Josep Gallifa,
Faculty of Psychology, Education and Sport Sciences Blanquerna, Ramon Llull University,
C. Cister 34, 08022 *Barcelona, Spain*
E-mail: josepgr@blanquerna.url.edu

DOI: https://doi.org/10.30564/jiep.v1i1.415

therefore, open to be answered and important is: Is there any pattern for modeling and understanding holistic consciousness?

In education, these tendencies helped to develop what has been known as holistic education. Holistic education is another way to refer the Early 20th. Century movement of "progressive education", carried out by Montessori, Steiner or Dewey, among many others. Their pedagogies emphasized imagination, aesthetics, organic thinking, practical engagement, creativity, and spirituality.[6] In the 70s diverse approaches emerged being "critical of the formal, modernist 'factory-model' of mass education. Most sought to broaden education beyond the simple information-processing model based on a mechanistic view of the human being to a more holistic, creative, multifaceted, embodied and participatory approach".[7] These perspectives were out of the dominant worldview of mainstream education. At the beginning of the 21st. Century appeared many pedagogical approaches related to the evolution of consciousness. Among them: aesthetic and artistic education, complexity in education, imaginative education, integral education, postmodern and poststructuralist pedagogies, social and emotional education, or spirituality in education.[8] These approaches, pedagogically well founded and structured, usually referred to the methods used as holistic pedagogies.

In a strict sense, holistic education designates some perspectives developed following a single pedagogue. It's the case, for instance, of Steiner's model of education, which "provides an integrated, holistic balance of intellectual/cognitive, artistic/imaginative and practical/life skills education".[9] This approach includes Steiner's notion of imaginative teaching using diverse methods: drama, exploration, storytelling, routine, arts, discussion, and empathy.[10] Another example, using similar methodologies, is Krishnamurti's holistic education, which "reminds us that we should not be dogmatic or doctrinaire in our education".[11] Some author-dependent holistic education developments become in time –unfortunately- "too narrow and rigid" [12] in their approaches.

In a more general sense, holistic education can designate any pedagogy that incorporates principles of spirituality, wholeness, and interconnectedness along with principles of freedom, autonomy, and democracy.[13] Doing so holistic education perspectives highlight eight broad principles: spirituality, reverence to life/nature, interconnectedness, human wholeness, individual uniqueness, caring relations, freedom/autonomy, and democracy.[14]

> "Holistic education is a radically non-reductionistic approach based upon a person-centered, ecological, global and spiritual worldview. As such, the holistic paradigm is an alternative not only to the scientistic reductionism of the modern age but also to the intellectual reductionism of postmodern thought. Holistic education is a humanistic as well as a spiritual critique of the dominant culture".[15]

Holistic education encompasses a wide range of pedagogies and philosophical or spiritual orientations. Holistic pedagogies try to include any significant aspect of the human experience, "the different aspects of the individual (intellectual, physical, spiritual, emotional, social and aesthetic), as well as the relationships between the individual and other people, the individual and natural environment, the inner- self of students and external world, emotion and reason, different discipline of knowledge and different form of knowing, holistic education is concerned with life experience, not with narrowly defined "basic skills".[16] Nevertheless, in holistic perspectives, sometimes there is an implicit -or sometimes explicit- worldview or metaphysical approach. A defining feature of holistic pedagogical perspectives is the relationship between education and consciousness evolution. This relationship has been promoted "pointing to the emergence of more complex, dialectical, imaginative, self-reflective and spiritual ways of thinking, living and loving".[17] Nevertheless, there isn't a unified systematic characterization of holistic consciousness, neither a characterization with minimal metaphysical references. This is the aim of this work: to find a general model for holistic consciousness to relate it to holistic education and pedagogies.

These aforementioned developments have relevance in today's general pedagogical debates, in a diversity of ways. For example, in the curricular field, there is a need to promote complex curricula, as opposed to the modernist oriented curriculum.[18] This means that nowadays there are new -or not so new- expected curricular standards: constructivist views of learning, social inclusion, and respect for diversity, a view of the process as being as important than product, lifelong learning outcome, the significance of metacognition and motivation in learning, or promoting multiple intelligences.[19] "Curriculum theory and practice are faced by new uncertainties, and such uncertainties require new approaches to practice and new ways of thinking".[20] One trend is framing of the new curricula around capacities or core/key competencies. Thus, for example, Scotland specifies that the curriculum should enable young people to become:[21] successful learners, confident individuals, responsible citizens, and effective contributors. Students are expected to play a diversity of roles in different dimensions of human evolution. And it means that there is a need for comprehensive curricular agendas. Another related pedagogical need is 'holding the

 DOI: https://doi.org/10.30564/jiep.v1i1.415

complexity', for example, using integrative and holistic approaches within the tradition of the critical-constructive Didaktik.[22]Yates defined 'pedagogy as the way we bring together broad intentions about what is to be learned or developed with particular kinds of learners in particular historical, national and institutional settings'.[23] The issues that have made this hard to deal with, in present times, are:[24]

1. The problem of reconciling recognition and distribution in relation to our greater sensitivity to diversity and embodied learners.
2. The problem of reconciling concepts of education based on comparative measurement and standards that look backward, with concepts of new times and a desire for young people to enter the changing world with a new kind of preparation and persona.
3. The continued divergence of cognitive and culturalist approaches to pedagogy and knowledge and identity.

To approach pedagogically these tensions, as well as other common contemporary pedagogical debates, it would be useful to have more well-defined models for holistic education.

The aim of this paper, to respond these contemporary needs, is: to present a theoretical framework for an ontology inclusive of the different levels of reality; to study, following a phenomenological path, a general model for holistic acts of consciousness, independent of metaphysical frameworks; to apply this model to education in order to characterize holistic education, and thus, to project these findings to develop holistic pedagogies.

2. Holonic Theory

Holonic theory can be the framework for an inclusive ontology. The holonic theory, based in the concept of holon that Koestler[25] coined, presents a model to integrate into a comprehensive model the evolution in the physiosphere, biosphere, and noosphere, terms that Teilhard de Chardin[26] developed in his understanding of evolution. The holonic model is useful to represent holistic consciousness acts, in particular, the full evolutionary spectrum of consciousness.[27] Some authors contributed to the development of the holonic theory. Among others, we can mention Goddard,[28] Smith,[29] Edwards,[30] Helfrich,[31] and Anderson.[32]

Minimal reference to Metaphysics

Wilber proposed a minimal Metaphysics at the moment of integrating the evolution from the physiosphere, towards the biosphere and finally towards the human noosphere,[33] in a similar way as Teilhard de Chardin did.[34] If this evolution is pushed by the Eros/will, there is another aspect, which is the result of the 'involution', that is to go back in the evolution, before the 'starting point'. This leads to the only metaphysical assumption that Wilber contemplates: the 'Agape' or genuine 'Love' in the beginning. Except for this presuppose, research, said Wilber, should be influenced as little as possible by metaphysical elements.[35] "With involution and evolution the basic idea is that Spirit as causal Consciousness 'steps down' or 'throws itself out' into soul, then mind, then body, and finally matter causing a Big Bang of primordial physios, which contains varying degrees of proto-interiority that sets the stage for matter to begin its journey back to Spirit through the evolutionary process (e.g., physiospheres, then biospheres, then noospheres, then...). Thus, the notion of involutionary and evolutionary 'currents', 'fields', or telos helps explain the plausibility that Spirit is hardwired into all physiospheres 'before the beginning', at the beginning (Big Bang), and thereafter providing a 'push and pull' or innate directionality and intention to the Kosmos".[36] Wilber in this regard stated:

> "Now you are perfectly free to believe in evolution and reject the notion of involution. I find that an incoherent position; nonetheless, you can still embrace everything… about the evolution of culture and consciousness, and refuse or remain agnostic about involution. But the notion of a prior involutionary force does much to help with the otherwise impenetrable puzzles of Darwinian evolution, which has tried, ever-so-unsuccessfully, to explain why dirt would get right up and eventually start writing poetry".[37]

A Definition of Holon

After presenting briefly the involution-evolution rationale, Wilber introduced the concept of holon that previously Koestler formulated:[38] A holon is a unit of consciousness that is at the same time part and whole. For example, an atom, a grain of sand, an animal, a symbol, are examples of holons. Holons can be considered themselves by definition as a whole and at the same time as part of another wider reality. Holons help to explain the hierarchical and heterarchical relationships within the evolutionary spectrum of consciousness.[39] Holons are the result of the interactions between any kind of potential realities with human consciousness.

Holons are organized and can be organizing agents. On the one hand "hierarchies unfold in irreversible, increasing depths of transcending yet include relationships, for instance: seedlings to saplings to trees. Heterarchies, on the other hand, consist of similar holons with spans of equivalence, for example, groups of atoms, amphibians, or primates".[40] Therefore the holon eventually would provide conceptual means to integrate crucial relationships during long periods of evolutionary time (evolution) with relationships more localized in a human life (development).

DOI: https://doi.org/10.30564/jiep.v1i1.415

Wilber refined Koestler's concepts of 'vertical' transformation and depth within 'horizontal' translation and span, two complementary aspects of development. Taken together, hierarchy (depths of inequivalence) and heterarchy (spans of equivalence) form what Koestler named a holarchy'.[41] Consequently, there are holons (whole/part) that can be: physical (ex: atoms), biological (ex: cells), psychological (ex: minds), psychic (ex: souls) and spiritual (ex. causal holons), and can be traced up and down in the holarchy.

Applying the definition of holon, "reality is not composed of things or processes, it is not composed of atoms or quarks. It is not composed of 'wholes', nor does it have any 'parts'. Rather, it is composed of whole/part or holon units".[42] "There is nothing that isn't a holon (upwardly and downwardly, forever".[43] With this definition, Wilber goes beyond atomism (reductionism, no interiors) and simple holism (extreme heterarchy). This moves him away from reductionism or incomplete synthesis. It must be taken into account that holons in this way defined comprise any simple or complex unit of consciousness.

A holon is, therefore, a fundamental structural element common to all reality. Its definition as a whole or as a part will depend on the context. It is always complete and incomplete so, trying to maintain its identity or express its potential, it flows and extends into becoming. It could be said that a holon is a construct, but also it is a self-evident revelation, resulting from a natural phenomenology. "Its symmetry provides structure and stability and its lack of definition provides asymmetries and movement. It is the 'root' event and the structural basis of the forms and of all the other events".[44]

Value of the Holons

Holonic theory served Wilber to propose a post-Metaphysics to understand the reality as a manifestation of the non-dual Spirit. The Spirit, whether non-dual or dual, as a basis and as an objective of the evolution, permeates both vertical holarchies and horizontal heterarchies. In this sense Wilber asserted that all holons have some kind of value:[45]

- Basic value: All holons are manifestations of the Spirit. This value is radically equal for all holons.

- Intrinsic value: The value that a holon has in itself as a whole (agent, depth in whole/part relationships).

- Extrinsic value: The value that the holon has as a part (communion, span in whole/part relationships).

"The higher the holon's vertical development within the Kosmos (e.g., physiosphere, biosphere, noosphere) the more intrinsic value and depth it has. On the other hand, the lower the holon's vertical development the more extrinsic value and greater span it has. Therefore, humans have more intrinsic value than animals, fish, prokaryotes, rocks, molecules, and quantum fields respectively, and vice versa for extrinsic value".[46] Nevertheless, these values are finally noospherical constructions or human consciousness constructions.[47]

Principles and Proprieties of the Holons.

Holons explain the 'relatively stable habits of evolution', rather than 'laws of evolution'; thus Wilber speaks of 'patterns of existence', 'results of the process of involution', 'tendencies of evolution', 'laws of form', 'propensities of manifestation'.[48] They have in common: "Eros: Each holon is Spirit-in-Itself playing at being 'other', because it is, by involution, another and has a tendency to evolve towards the Spirit; Agape: The Spirit reaches out to all holons attracting other holons; Morphogenetic gradient or field of potentials: result of involution with contents and concrete forms but with evolutionary potential; and also certain fixed patterns or prototypical forms".[49]

Holons manifest in 20 tenets. These are:[50]

1. Reality as a whole is not composed of things or processes but of holons (part/whole).{1}
2. Holons display four fundamental capacities: self-preservation, self-adaptation, self-transcendence. self-dissolution.{2}
3. Holons emerge, because transcendence, freedom, and creativity are consubstantial in them.{3}
4. Holons emerge holarchically (as whole/part).{4}
5. Each emerging holon transcends and includes its predecessor(s).{5}
6. The lower or less inclusive holon sets the possibilities of the superior or more inclusive. The higher or more inclusive holon sets the probabilities of the lower or less inclusive.{6}
7. The number of levels, which a hierarchy comprises, determines whether it is shallow or deep (vertical dimension); {7} and the number of holons on any given level we shall call its span (horizontal dimension).{8}
8. Each successive level of evolution produces greater depth and less span. Addition I: The greater the depth of a holon, the greater its degree of consciousness.{9}
9. Destroy any type of holon and you will destroy all the holons above it and none of the holons below it.{10}
10. Holarchies coevolve.{11}
11. The micro (individual holons) is in relational exchange with the macro (social/ environmental holons) and all levels of its depth.{12}
12. Evolution has directionality,{13} increasing complexi-ty,{14} increasing differentiation/integration, {15} increasing organization/structure,{16} increasing relative autonomy. {17} Increasing 'telos'.{18}

 DOI: https://doi.org/10.30564/jiep.v1i1.415

13. Addition II: Every holon issues an IOU to the Cosmos (IOU = Incomplete Or Uncertain). In other words, it is in tension and constant search.{19}
14. Addition III: All IOUs to the Cosmos are redeemed in Emptiness. Holons are conventional truths. The Emptiness/Spirit is the ultimate truth.{20}

These principles describe the nature of the holons and their behavior in the framework of the process of 'involution-evolution'. A central aspect is the inclusion of "telos", directionality, purpose, which Wilber takes from Withehead, as well as he did with the involution-evolution model. In the words of Wilber:

> "Whitehead had the notion that reality appears moment by moment as a subject of experience, and even the smallest material objects, such as molecules or atoms, have a small level of consciousness; a little proto-feeling that he called "apprehension." In this way, this present moment is a subject of experience that apprehends the previous moments as an object, feeling the experience as it appears from moment to moment. So the previous moment, which was a subject before, now becomes the 'object' of the subject. And that is causality because the past is having an impact on the present, determining it. If this were the only thing that could happen (the subject that apprehends the previous object over and over again), the Universe would be deterministic, mechanistic and causal. But according to Whitehead every moment, besides having 'apprehension', has 'creativity', which means that there is a bit of originality or novelty".[51]

This evolutionary 'creativity' or 'novelty' can be explained in more detail: In every moment there is a bit of novelty. It is an optimization of the Universe. Every holon, even those of inner matter, has a slight degree of consciousness, and a small degree of creativity. Each moment transcends and includes the previous moment, adding novelty, since it goes beyond the previous. That's what is happening even at the lowest levels. Then there will never be a strict causality since there are causality and creativity

Self-transcendence explains the emergence of new holons or evolution. Therefore, "in the model, freedom and self-determination are foundational (Prigogine). The determinism arises when there is no self-transcendence or when is limited by a higher holon. When creativity is minimal, reconstructive sciences collapse in predictive sciences; this is the source of all reductionism".[52] This means that when transcendence is near zero the realist ontology and the predictive logical-empirical sciences are particular cases of the general model. But the general model encompasses other possible ontologies and epistemologies.

Thus evolution has directionality "exists in terms of increasing differentiation, variety, complexity, and organization in the physiosphere, biosphere, and noosphere. Regressions, dissolutions, or stoppages occur, but other indicators of directionality include creative emergence (novelty), symmetry breaks (Prigogine), self-transcendence, increased depth, and greater awareness".[53] Thus, there is greater complexity, greater differentiation/integration, increasing in organization/structure, increasing in relative autonomy in the midst of social/environmental fluctuations, the realization of 'telos'.

This last aspect of 'telos' assumes that each holon also acts as an attractor, an 'omega point', for the actualization of itself or other holons in space and time. 'Pulls' the updating or development of the holon in one direction, be it physical, biological or mental system.[54] Broader and wider contexts 'pull', in the form of 'telos', more limited present contexts. However, in the biosphere, living holons also do other things such as sexual reproduction, metabolic communication, autopoietic self-preservation, etc. The holons with mind of the noosphere do things that biospheres cannot do, such as verbal communication, conceptual self-expression, artistic effort, etc. Thus the general laws are completed with these new realities.[55]

Detailed Dimensions of Holons

Wilber synthesized what holons of any kind have in common. Every holon has four characterizing dimensions:[56] Agency. A tendency to be a whole. Aristotelian entelechy, morphic unit/field (Sheldrake), canon (Koestler), self-asserting, relative autonomy and wholeness, yang. It manifests the tendency towards self-preservation, autonomy, self-responsibility, self-esteem. It assumes in this sense fixed forms or patterns, among which there are the 20 tenets. Wilber named this dimension deep structure. In pathological forms, it manifests as alienation and repression.

Communion. A tendency to the relationship, participatory, bonding, joining tendencies, expresses its partnership, the ability to be part of a whole, attract other parties, relationship with something larger, self-adaptation, yin. Pathological forms: fusion and indisociation.

Self-transcendence. Self-transformation, creative novelty, creativity.[57] Each holon becomes a new whole/part that has its own new forms of agency and communion. It is about the impulse to experience freedom, to find cohesion and unity through a greater, deeper and broader totality. Articulated by 'symmetry breaks' (Prigogine) not equivalent rearrangements of the same stuff. Evolution is the result of self-transcendence at all levels: It is also called as 'Eros', that is, Spirit manifested in something else: matter, body, mind, soul, etc. In this dimension, the 'telos' or purpose is manifested. If self-transcendence is

  DOI: https://doi.org/10.30564/jiep.v1i1.415

not achieved, 'phobos' (fear, regression, panic, contraction, and repression) is experienced.

Self-dissolution/Self-immanence. Self-dissolution of transcendence that can be termed as self-immanence. A morphogenetic gradient in the manifest field. This means not only a manifest reality with some kind of support in the manifested reality but also potential to evolve. Preservation of the current level or regression to previous levels. Wilber conceptualizes it as an instinct of death or Thanatos, a force opposed to Eros.

The four properties can be represented in axes, as Wilber proposed: "Taken together, these four capacities can be imagined as a cross: Two horizontal opposites: agency and communion, and two vertical opposites: self-transcendence and self-immanence".[58] We proposed, in previous work, a particular location for these dimensions in the space.[59] The simple representation of a holon is, therefore:

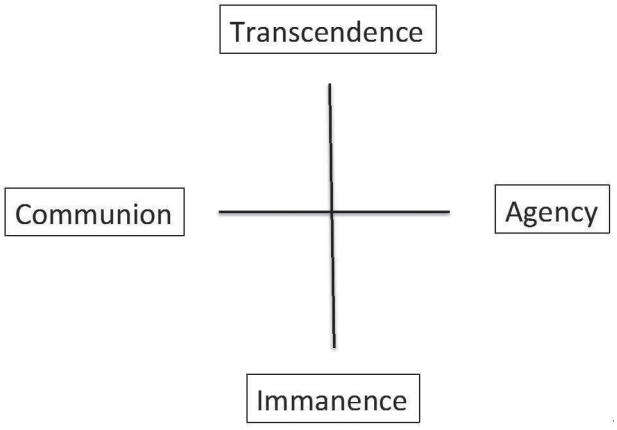

Figure 1. Immanence-Transcendence-Agency-Communion

Neurobiological Correlates

An example of holon can be a neuron, basic unit of the human brain: It is a cell and, as such, requires nutrients and oxygen to be alive and active, there is a somatic metabolism (preservation); it has a structure-nucleus, axon, etc., that is optimized by making connections and also by new connections that allow electrical or chemical synapses (agency); it is related to other neurons or groups of neurons - through connections in the dendrites - with which it forms patterns or neural networks (communion); it is open and sensitive to novelty in its environment (changes in neurotransmitters at synapses, new connections, new patterns) and activates/changes DNA to adapt to its environment and regulate its behavior over time (transcendence). The holonic scheme could be applied equally to more wide neuronal patterns or to the brain as a whole.

Introducing holons we do not necessarily assume any kind of unnoticed dualism because of the relationship that exists between mind and brain. Damasio[60] clarified the mind-brain relationship, understanding the relationship between neuronal patterns and mental images (equivalent to holons):

> "When I say that images depend and arise from neuronal patterns or neural maps, instead of saying that they are neural patterns or maps, it is not that I slide towards an inadvertent dualism, that is, a neuronal pattern on the one hand and a 'cogitum' immaterial by another. Quite simply, what I am saying is that we can not yet characterize all the biological phenomena that occur between: a) our current description of a neuronal pattern at various neuronal levels, and b) our experience of the image originated within the activity of the neuronal map. There is a gap between our knowledge of neuronal events in the molecular, cellular and systems fields, on the one hand, and the mental image whose appearance mechanisms we wish to understand. There is a vacuum that must be filled with physical phenomena not yet identified, but presumably identifiable. The size of that gap and the degree to which it can be more or less saved in the future is, of course, an object of debate".[61]

In other words, by introducing mental images Damasio doesn't separate his research from the neuropsychological paradigm but discarded the position of the "naive realism" by identifying, without further ado, neuronal patterns (physical, brain) and mental images (mind, for example, holons). Brain and mind are very closely related but they are not exactly the same.

3. Method: Phenomenological Psychology

To approach the proposed aims we used a phenomenological approach. Giorgi,[62] studying the whole person and not only fragmented psychological processes, developed a "phenomenological method for researching humans in a psychological way based upon the work of Husserl and Merleau-Ponty".[63] He explained that "psychologically phenomenologists are interested in specifically human consciousness, and because of that limited interest, the phenomenological method needs to be pre-transcendental",[64] which means that psychology is interested in how a human consciousness relates to a specifically human world.[65] Holons are the privileged location where this relationship takes place. Van den Berg emphasized that "insights into experience as lived, or the phenomenal level, is what is critical for psychologists to understand".[66] This is the level that corresponds to the modalities of consciousness and education, which we are interested in. So the phenomenological-psychological method is appropriate to our aims. Therefore our characterization followed

 DOI: https://doi.org/10.30564/jiep.v1i1.415

the phenomenological method. To accomplish "the criteria necessary in order for a qualitative scientific method to qualify itself as phenomenological in a descriptive Husserlian sense, one would have to employ (1) description (2) within the attitude of the phenomenological reduction, and (3) seek the most invariant meanings for a context".[67] These attitudes have been applied in our study.

4. Results

A Model for Holistic Acts of Consciousness

The holonic theory purpose is to provide a unified theory of the evolution, inclusive of the evolution of consciousness. Wilber took care especially about the basics, giving a coherent explanation of the evolution of the physiosphere and biosphere, but giving no special details about the noosphere. To accomplish our purpose, which is to find a model for holistic acts of consciousness, we'll study phenomenologically diverse noospheric holons, covering diverse levels. The selected holons are either different dimensions of a whole, or complete systems to explain consciousness, in both cases represented in their holonic structure. This table covers a wide range of dimensions of human faculties. After studying carefully each one of the holonic four dimensions we'll give a name to each one. A general holon will rise above the included holons.

After considering and studying the diverse dimensions, the last row has been defined with names that characterize in each case the whole corresponding column. A general and inclusive holon emerge in this last row. Because it's a holon inclusive of the diverse noospheric holons and more elevated can be a representation of a holistic act of consciousness. The dimensions of this emergent holon include:

Preservation. It's the dimension of immanence (I). It's the preservation of what has value, frequently it comes from the past. It can have positive aspects, for example in resilience.

Profundity. It's the dimension of transcendence (T). More profundity means more consciousness (tenet number {9}). Without that dimension, there wouldn't be an advance to-wards holistic consciousness.

Projective action. It's the dimension of agency (A). The internal structure and order can be projected outside. Acting is a way of exercising free will. Although limited, there is room for human freedom. Action can be interiorized in a thinking process.

Span. It's the dimension of communion (C). This means participatory bonds and mutual understanding through shared meanings and implicit values.

In a holistic act of consciousness, these dimensions have to be, in some way, present. Note for example that in a case that there had been span but not profundity, there wouldn't have been the implication of a fundamental dimension for the advancement of consciousness. The consideration of this case as an act of holistic consciousness would be problematical. And the same can be said with the other dimensions.

A Model for Holistic Education

Once we have a model to include holistic acts of consciousness we can apply this holistic form of consciousness to education. Doing so we'll develop a model to characterize holistic education. Therefore, holistic education, following the development made, needs the contribution of four rationales:

Preservation (I). There are valuable things that must

Table 1 Noospheric Human Holons[68]

Human holons:	Immanence	Agency	Communion	Transcendence
Organismic self-regulation (instincts)	Conservation	Sexual	Social	Novelty
Categories of perception	Time (Kronos)	Movement	Space	Time (Kairos)
Naranjo: Character-Essence	Dominant passions	Fixations	Virtues	Holy ideas
Enneagram dominant passions #	8, 9, 1	6, 7	2, 3	4, 5
Tolle: Ego-Being	Pain-body Past	Mind Past/Future	Presence Now	Being Unmanifested
Levels of desire	Receiving for oneself alone (ego)	Receiving	Receiving for the sake of sharing	Sharing (soul)
Human faculties	Memory	Will	Understanding	Wisdom
Cardinal virtues	Fortitude	Temperance	Prudence	Justice
Wilber's quadrants	Objective	Interobjective	Intersubjective	Subjective
Laloux: styles of Organization	Red-amber	Orange	Green	Teal
General holon	Preservation	Projective action	Span	Profundity

DOI: https://doi.org/10.30564/jiep.v1i1.415

be preserved in education. For example: the already established knowledge via diverse sciences. On the other hand preservation in education, in a more concrete way, takes the form of a program. A program has objectives, methods, evaluation procedures, etc. It can have different dimensions and levels. Can be an administrative, strategic, curricular or classroom-applied program. Preservation means to preserve all the important values even the ones that come from the tradition. The recent trend to implement evidence-based reforms[69] can be an example that confirms the importance of enhancing education with the implementation of proved programs. There are programs that have to be maintained. It's the logic of resources and energies to maintaining valuable educational efforts. That's the first rationale.

Profundity (T). The whole human being has to be considered. This is the second rationale. Education can be understood as a mean to improve consciousness and as a way of promoting human growing. It can be also promoting the development of the human faculties in developmental lines.[70] Particularly interesting is the progress toward advanced stages of consciousness.[71] In this way, education of the whole human being can be understood as a part of a wider general movement of elevation of personal consciousness.[72] Psychologically speaking can be the construction of a bridge between the unconscious and conscience as in the Jungian transcendent function.[73] It is also the rationale of the premodern religious-spiritual traditions.

Projective action (A). That means the knowledge of a group in order to transform reality by constructing artifacts.[74] This means including the anthropological rationale, the ways mean by which the humanity had developed. It implies intentionality, the election of means, transformative action, evaluation, as well as learning in an expert-novice framework.[75] Metacognition acquisition and competences based education fall inside that modality. A correspondence can be made with the Aristotelian episteme poietike directed to the poiesis. The educative professions are also in evolution in this category. It's the tékhne-arts-creative-productive logics.[76] This means introducing practical and productive life and its rationale to education.

Span (C). This is the evolutionary cultural dimension. Culture is part of the imagined representations.[77] This dimension is structured in worldviews, which are frequently unconscious until the advanced stages of consciousness. The evolution of the different worldviews and the values or v-memes of a culture was studied by Beck & Cowan. [78] This is the logic of the Aristotelian episteme praktike oriented to the phronesis. It's the place for values and ethics. Cooperative learning and team learning are located in this dimension. This is another rationale to be taken into account.

5. Discussion

Holonic Theory

We defined and developed the holonic theory as an ontological theory to explain the different levels of reality. There are holons (whole/part) in the physiosphere, in the biosphere, and in the noosphere. Examples of holons are a simple neuron, a neuronal pattern, a brain or a human being. Holonic theory is useful to integrate science, arts, cultures, and levels of consciousness. This perspective is not necessarily dualist (Damasio).

A correspondence can be done between scientific rationale and the Aristotelian episteme theoretike, conducted by the logical necessity. The corresponding Aristotelian way to the profundity of consciousness is the Aristotelian nous. And episteme poietike and praktike correspond respectively to projective action and culture. The predominance of each holonic dimension defines different epistemological approaches: empiric-logical science (I), phenomenological (T), Aristotelian tradition (A) and constructivist (C).[79] Additionally, the model provides the background necessary to support Wilber's integral theory.

Holistic Consciousness

We studied diverse noospheric holons and found four dimensions useful in the characterization of holistic acts of consciousness: Preservation, profundity, projective action and span, We can understand from the model developed why only growing in one dimension is not fully holistic. This can be applied to the common contemporary syncretism, frequently denominated as holistic, that has span but without concern in profundity. The same can be said about some consciousness development movements that search for profundity but maybe discarding other logics, like for example preserving the good of the tradition. Truly holism, in harmony with the holonic theory, needs to contemplate the four dimensions in order to facilitate the integral evolution of consciousness.

Holistic Education

Following the model developed we defined holistic education, a way to consider education through four rationales: program, whole-person, projective action, and culture. Our suggestion for promoting a truly holistic education is to study how it can be referred to this characterization presented. We can see the limitations of the educational 'holisms' that are only activism, the ones only content-based, the ones focused only in personal development or the proposals that are only a juxtaposition of different perspec-

 DOI: https://doi.org/10.30564/jiep.v1i1.415

tives. In order to develop a model of holistic education,[80] a reasonable presence of the four dimensions seems appropriate, after the development made by applying the holonic theory and the holistic consciousness to education. When in addition the holistic consciousness uses the whole holarchy (Wilber's integral theory) the resulting approach has been denominated as integral education.[81]

Application into the Practical Context

We'll present a practical example for Project Based Learning, a usual methodology widely applied in education. Holistic education would be to project in this methodology the consciousness of the dimensions. Applying the development described, any educational activity has an impact in the four dimensions presented:

- Program. To select a formative activity with pieces of evidence in the past on their educative value. Value related to curricular contents and/or with past successful experiences.
- Whole-person. In which way this activity will help in the advance of consciousness of participants. Will it be relevant to the participant's lives?
- Projective action. How this activity will help in transforming reality. What kind of competences and abilities will be shared and mastered?
- Culture. What are the implicit values in the activity? Are there opportunities to advance toward collaborative and dialogical values?

Each one of the dimensions has their own rationality. Each one is important in a holistic education activity and can be assessed independently. In our example is not only important to solving the problem, but also the educational treatment of the other dimensions. Holistic education means to be conscious, or more precisely to apply holistic consciousness, into educational situations. This can be programmed and evaluated in a holistic pedagogy.

6. Conclusion

Begging with the detection of the contemporary need for holistic consciousness and integral visions, we developed the holonic theory, and characterized a holistic act of consciousness, using a psychological phenomenological method.

We related the many times blurry defined holistic education with holistic consciousness. We characterized holistic education as a holistic consciousness act with four dimensions: program, whole-person, projective action, and culture. All of these rationales are implicated in any holistic education act. Sometimes in regular education, certain of these dimensions remain unconscious. In this case, the perspective under consideration may be educative but not holistic or product of a holistic consciousness act.

All of these developments can be applied to pedagogy: The characterized holistic education can serve as a systematic framework for holistic pedagogical developments and debates. Holistic consciousness allows holistic pedagogical discussions to go beyond modern worldview and postmodern syncretism.

References

[1] Wilber K. The theory of everything. Boston, MA: Shambhala; 2001.
Wilber K. Introduction to Integral Theory and Practice IOS basic and the AQAL map. AQAL Journal of Integral Theory and Practice 2005; 1(1): 1-38.
Wilber K. The integral vision: A very short introduction. Boston, MA: Shambhala; 2007.

[2] Beck, DE., Cowan, C. Spiral Dynamics: Mastering Values, Leadership, and Change. New York: Wiley/Blackwell; 2005.

[3] Esbjörn-Hargens S. Integral Education By Design: How Integral Theory Informs Teaching, Learning, and Curriculum in a Graduate Program. ReVision 2005; 28(3): 21-29.
Murray T. What is the Integral in Integral Education? From Progressive Pedagogy to Integral Pedagogy. Integral Review 2009; 5(1): 1-38.

[4] Laloux F. Reinventing Organizations: A Guide to Creating Organizations Inspired by the Next Stage of Human Consciousness. Brussels: Nelson Parker; 2014.

[5] Thomas B. Integral Leadership Manifesto. [Internet]; 2018 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from: http://integralleadershipmanifesto.com .

[6] Gidley JM. Evolution of education: From weak signals to rich imaginaries of educational futures. Futures 2012; 44(1): 49.

[7] Gidley JM. op. cit.; 2012: 50.

[8] Gidley JM. op. cit.; 2012: 50.

[9] Gidley JM. Holistic education and visions of rehumanized futures. RoSE–Research on Steiner Education 2010; 1(2): 139.

[10] Nielsen TW. Towards a pedagogy of imagination: A phenomenological case study of holistic education. Ethnography and Education 2006; 1(2): 247-264.

[11] Miller JP. Krishnamurti and holistic education. Encounter, Education for Meaning and Social Justice 2000; 13(4): 42.

[12] Miller JP. Krishnamurti and holistic education. Encounter, Education for Meaning and Social Justice 2000; 13(4): 43.

[13] Rudge LT. Holistic education: An analysis of its pedagogical application (Doctoral dissertation, The Ohio State University); 2008. p 2.

[14] Rudge LT. Holistic education: An analysis of its pedagogical application (Doctoral dissertation, The Ohio State University); 2008. p 3.

  DOI: https://doi.org/10.30564/jiep.v1i1.415

[15] Miller R. Beyond reductionism: The emerging holistic paradigm in education. The Humanistic Psychologist 2000; 28(1-3): 382.

[16] Mahmoudi S., Jafari E., Nasrabadi HA., Liaghatdar MJ. Holistic Education: An Approach for 21 Century. International Education Studies 2012; 5(2): 178.

[17] Gidley JM. Educational imperatives of the evolution of consciousness: The integral visions of Rudolf Steiner and Ken Wilber. International Journal of Children's Spirituality 2007; 12(2): 117-135.

[18] Morrison K. Complexity theory and curriculum reforms in Hong Kong. Pedagogy, Culture and Society 2003; 11(2): 279-302.

[19] Morrison K. Complexity theory and curriculum reforms in Hong Kong. Pedagogy, Culture and Society 2003; 11(2): 279-302.

[20] Priestley M. Whatever happened to curriculum theory? Critical realism and curriculum change. Pedagogy, Culture & Society 2011; 19(2): 223.

[21] Priestley M. Whatever happened to curriculum theory? Critical realism and curriculum change. Pedagogy, Culture & Society 2011; 19(2): 221-237.

[22] Hudson B. Approaching educational research from the tradition of critical-constructive Didaktik. Pedagogy, Culture and Society 2003; 11(2): 173-187.

[23] Yates L. From curriculum to pedagogy and back again: Knowledge, the person and the changing world. Pedagogy, Culture & Society 2009; 17(1): 27.

[24] Yates L. From curriculum to pedagogy and back again: Knowledge, the person and the changing world. Pedagogy, Culture & Society 2009; 17(1): 27.

[25] Koestler A. The ghost in the machine. New York, NY: MacMillan; 1967.

[26] Teilhard de Chardin P. The phenomenon of man. New York NY: Harper Collins; 2002.

[27] Wilber K. Spectrum of consciousness. Wheaton, IL: Quest; 1993.

[28] Goddard G. Holonic logic and the dialectics of consciousness. Unpacking Ken Wilber's four quadrant model [Internet]; 2000 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from:http://www.integralworld.net/goddard2.html.

[29] Smith AP. The Spectrum of holons. A response to Fr Kofman. [Internet]; 2001 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from: www.integralworld.net/smith2.html.

[30] Edwards M. A brief history of holons [Internet]; 2003 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from: http://www.integralworld.net/edwards13.

[31] Helfrich PM. Ken Wilber's model of human development. An overview. [Internet]; 2007 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from: http://www.paulhelfrich.com/library/Helfrich_P_The_Five_Phases_of_Wilber.pdf
Helfrich PM. Ken Wilber's AQAL metatheory: An overview. Castaic, CA: Wildfire Media; 2008.

[32] Anderson P. A fully holonic meta-theory for the heart of integral. [Internet]; 2017 [updated 2018 Dec 10; cited 2018 Dec 16].Available from: http://www.integralworld.net/philanderson1.html.

[33] Wilber K. Up from Eden: A transpersonal view of human evolution. Wheaton, IL: Quest; 1996.

[34] Teilhard de Chardin P. The phenomenon of man. New York NY: Harper Collins; 2002.

[35] Helfrich PM. op. cit.; 2008.

[36] Helfrich PM. op. cit.; 2007.

[37] Wilber K. Introduction to the collected works of Ken Wilber: Volume 2. [Internet]; 2003 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from: http://wilber.shambhala.com/html/books/cowokev7_intro.cfm/.

[38] Koestler A. The ghost in the machine. New York, NY: MacMillan; 1967.

[39] Helfrich PM. op. cit.; 2007.

[40] Helfrich PM. op. cit.; 2007. p 3.

[41] Helfrich PM. op. cit.; 2007.

[42] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 41.

[43] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 41.

[44] Piazenza G. El Holón y la dualidad inicial de la complejidad (Spanish) [Internet]; 2012 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from: https://www.monografias.com/trabajos94/holon-dualidad-inicial-complejidad/holon-dualidad-inicial-complejidad.shtml. p 1.

[45] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 544.

[46] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 545.

[47] Helfrich PM. op. cit.; 2008.

[48] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 40.

[49] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 132-134.

[50] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 43-85 (tenets 1-18) and 527-532 (tenets 19 and 20).

[51] Wilber K. Ken Wilber entrevistado por Raquel Torrent en 2016. Con preguntas de la Comunidad Integral Europea (Spanish) [Internet]; 2016 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from:https://asociacionintegral.

 DOI: https://doi.org/10.30564/jiep.v1i1.415

wordpress.com/. p 18.

[52] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 55.

[53] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 74.

[54] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 81.

[55] Wilber K. Sex, ecology and spirituality. Boston, MA: Shambhala; 1995. p 121-124.

[56] Wilber K. op. cit.; 1995. pp 48-52 and 132-134.

[57] Whitehead AN. Process and Reality. New York: The Free Press; 1978.

[58] Wilber K. op. cit.; 1995. p 54.

[59] Gallifa J. Paradigms and Methodologies for Knowledge Building. Review of Educational Theory 2018b; 1(3): 73.

[60] Damasio AR. La sensación de lo que ocurre. Cuerpo y emoción en la construcción de la conciencia. (Spanish). Madrid: Editorial Debate; 2001.

[61] Damasio AR. op. cit; 2001. p 327.

[62] Giorgi A. The Descriptive Phenomenological Psychological Method. Journal of Phenomenological Psychology 2012; 43(1).

[63] Giorgi A. The Descriptive Phenomenological Psychological Method. Journal of Phenomenological Psychology 2012; 43(1): 3.

[64] Giorgi A. The Descriptive Phenomenological Psychological Method. Journal of Phenomenological Psychology 2012; 43(1): 9.

[65] Giorgi A. The Descriptive Phenomenological Psychological Method. Journal of Phenomenological Psychology 2012; 43(1): 10.

[66] Giorgi A. The Phenomenological Psychology of J.H. van den Berg. Journal of Phenomenological Psychology2015; 46(2): 141.

[67] Giorgi A. The Theory, Practice, and Evaluation of the Phenomenological Method as a Qualitative Research Procedure. Journal of Phenomenological Psychology 1997; 28(2): 235.

[68] Naranjo C. Carácter y neurosis. Una visión integradora (Spanish). Vitoria: La llave; 1996.
Tolle E. The power of now: A guide to spiritual enlightenment. San Francisco: New World Library; 2004.
Berg Y. El poder de la Kabbalah (Spanish). Los Angeles CA: Centro de Càbala Internacional; 2006.
Wilber K. The theory of everything. Boston, MA: Shambhala; 2001.
Wilber K. Introduction to Integral Theory and Practice IOS basic and the AQAL map. AQAL Journal of Integral Theory and Practice 2005; 1(1): 1-38.
Wilber K. The integral vision: A very short introduction. Boston, MA: Shambhala; 2007.
Laloux F. Reinventing Organizations: A Guide to Creating Organizations Inspired by the Next Stage of Human Consciousness. Brussels: Nelson Parker; 2014.

[69] Slavin RE. Evidence-based education policies: Transforming educational practice and research. Educational researcher 2002; 31(7): 15-21.

[70] Gardner H. Frames of mind: the theory of multiple intelligences. New York: Basic Books; 1983.

[71] Botella L., Gallifa J. A constructivist approach to the development of personal epistemic assumptions and worldviews. Journal of Constructivist Psychology 1995; 8(1): 1-18.
Commons ML., Richards, FA. Four postformal stages In: Demick J (editor) Handbook of adult development. New York: Plenum Press; 2002. pp. 199-219.

[72] Graves CW. Human Nature Prepares for a Momentous Leap. The Futurist; 1974: 72-87. [Internet]; 1974 [updated 2018 Dec 10; cited 2018 Dec 16]. Available from: http://www.clarewgraves.com/articles_content/1974_Futurist/1974_Futurist.html.
Wilber K. Introduction to Integral Theory and Practice IOS basic and the AQAL map. AQAL Journal of Integral Theory and Practice 2005; 1(1): 1-38.

[73] Jung CG. Analítical psychology: its theory and practice: the Tavistock lectures. New York: Vintage books; 1968.
Jung CG. The basic writings of C.G. Jung. New York: The Modern Library; 1993.

[74] Helfrich PM. op. cit.; 2008.

[75] Gallifa J. The Tekhne-logic revolution. Rethinking the 'interobjective' dimension of the Integral Theory. Consequences and relevance to Education. Creative Education 2018; 9(7): 1084-1104.

[76] Gallifa J. op. cit.; 2018: 1084-1104.

[77] Harari YN. Sapiens: A Brief History of Humankind. New York: Random House; 2014.

[78] Beck DE., Cowan C. Spiral Dynamics: Mastering Values, Leadership, and Change. New York: Wiley/Blackwell; 2005.

[79] Gallifa J. Paradigms and Methodologies for Knowledge Building. Review of Educational Theory 2018b; 1(3): 70-81.

[80] Forbes S. Holistic education: An analysis of its ideas and nature. Brandon VT: Foundation for Educational Renewal Publishers; 2003.

[81] Esbjörn-Hargens S. Integral Education By Design: How Integral Theory Informs Teaching, Learning, and Curriculum in a Graduate Program. ReVision 2005; 28(3): 21-29.
Murray T. What is the Integral in Integral Education? From Progressive Pedagogy to Integral Pedagogy. Integral Review 2009; 5(1): 1-38.
Gallifa J. Educació integral: transformar l'educació, formar a la persona, millorar el món (Catalan). Barcelona: Editorial UOC; 2018c.

  DOI: https://doi.org/10.30564/jiep.v1i1.415

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep/

ARTICLE

Example of a Project to Work the Sustainability Competence in the Subject of Linear Algebra in Engineering Studies

M.I. Garcia-Planas* and J. Taberna
Universitat Politècnica de Catalunya, Departament de Matemàtiques, Barcelona, Spain

ARTICLE INFO

Article history:
Received: 12 November 2018
Accepted: 16 January 2019
Published: 18 January 2019

Keywords:
Sustainability
Project-based-learning
Linear algebra

ABSTRACT

Strength in sustainability is becoming more and more essential for anyone who wants to dedicate themselves to the practice of competent and professional engineering. From the scientific-technical teachings, therefore, the transversal competencies of sustainability and social commitment should be addressed, so that students understand and become aware of the problems of a global world. Using the project-based learning tool (PBL), from a practical case, transversal competencies related to the global issues of the Sustainable Development Objectives (SDG) can be explained and evaluated in a classroom of Engineering or scientific degrees through the subject of Linear Algebra. This project presented is novel since usually, in a class of mathematics, teacher proposes examples in which we can see an attempt to approximate math to essential topics such as sustainability. In this case, the procedure is the other way round; it is about setting as the objective of the course the solution of a problem related to sustainability and adapting the mathematics program so that throughout the course the problem is solved. It is not about looking for application examples of the tools, but looking for algebra tools to solve the problem.

1. Introduction

In 2015, was adopted the 2030 Agenda for Sustainable Development for transform the world and 17 Sustainable Development Goals were fixed. One of the goals is the number four where in its target 4.7 ensures (see Wells[19]), that all learners acquire the knowledge and skills needed to promote sustainable development including education, global citizenship and cultural diversity. On the other hand, all the objectives admit a mathematical approach that allows them to advance in the achievement of the same. Therefore, it is important to introduce these aspects in higher education and the mathematics subjects so that students become aware of them and at the same time, they observe that the knowledge of mathematics contributes positively in the achievement of the Objectives of Sustainable Development (SDG).

**Corresponding Author:*
M. I. Garcia-Planas,
Universitat Politècnica de Catalunya, Departament de Matemàtiques, Barcelona, Spain;
Email: skhawaldeh@aabu.edu.jo

 DOI: https://doi.org/10.30564/jiep.v1i1.317

The University has a basic function to play in implementing the 2030 Agenda for Sustainable Development. It is an essential performer in achieving quality education, but it must also contribute to the training of responsible citizens committed to the local sustainable development and global.

Progressively, universities and other institutions of higher education have been incorporating values and practices of sustainable development in its core teaching and research activities, institutional management and operational systems, even some of them have prepared strategy guides to integrate the learning of the generic competition "sustainability and social commitment" [21]. However, so far they have mainly focused on justifying and reasoning about the need to implement competencies on sustainable development but without taking action.

To break this dynamic and choose one of the objectives of sustainable development, we introduce in an evaluable way the competence of sustainability and social commitment in a subject of the first year of studies of the degree of engineering, in order that from the beginning, students take Awareness of the global problem of sustainable development.

It is clear that having a dedicated course on sustainability does not guarantee the students' full understanding of the sustainable message. Therefore, a solution as already indicated Colombo et al.[4], Allen et al.[1] and Murphy et al.[14] could be to transpose transversely into an existing course unit the sustainability elements on which it is desired to focus.

From the scientific-technical teachings, the transversal competencies of sustainability and social commitment must be addressed, so that students understand and become aware of global problems. Using the project-based learning tool (PBL), from a practical case, transversal competencies related to the global issues of the Sustainable Development Objectives[23], can be explained and evaluated in the classroom.

Crofton[6] indicates that provide education based only on technical knowledge is insufficient to address the complexity of the problems associated with sustainable development. It is also noteworthy that there is a growing demand from students to rethink the content and form of courses taught in engineering degrees, (Trimingham et al.[18]). To deal with all this requires training that helps to contemplate environmental problems and development as a whole.

Some essays over sustainability competence implementation in engineering studies have been reported in recent years, for example, an interdisciplinary Project Based Learning strategy has been applied in the first year of the Integrated Master's degree in Industrial Engineering and Management at University of Minho (Colombo et al.[5]). Nevertheless, there are no essays in which the only subject is Linear Algebra.

This paper presents a proposal of a project to implement the competence of sustainability and social commitment through PBL in the Linear Algebra subject of the first course of the engineering curriculum at any Engineering schools or Sciences faculties that contain this matter in its curriculum.

With the proposed implementation of Project-based Learning, we not only introduce the competition corresponding to the sustainable development objective of a specific goal (in this case, the number 6, although it is possible to prepare projects for each of the objectives (see Garcia et al.[9]), but also address the number 4 one that is related to education. This target is to "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all". Given that the PBL promotes quality education and supports curricular innovation, it collaborates in consecution of the SDG number 4.

The project presented is novel since usually, in a class of linear algebra; the teacher proposes examples in which we can see an attempt to approximate linear algebra to essential topics such as sustainability but always from mathematical matter. In this case, the procedure is the other way round; it is about setting as the objective of the course the solution of a problem related to sustainability and adapting the linear algebra program so that throughout the course the problem is solved. It is not about looking for application examples of the tools, but looking for algebra tools to solve the problem.

This work is not about introducing the PBL as a teaching method of linear algebra, but to show that its use can be a good method to evaluate the transversal competence of sustainability and social commitment through a STEM subject as well as to help students to reflect on the SDGs.

Integration of sustainability into the curricula of engineering education

Since the World Commission on Environment and Development had presented the "Our Common Future" known as the Brundtland Report[20], to the General Assembly in 1987, the concept of sustainable development has gained increasing interest and attention. Sustainable development is defined as development that meets the needs of the current generation without compromising the ability of future generations. However, the lack of understanding and even misinterpretation of this definition has delayed the implementation of sustainable practices within the framework of Higher Education.

Universities must act as agents of change by promot-

  DOI: https://doi.org/10.30564/jiep.v1i1.317

ing the principles of sustainable development within their institutions and in society. As Barth and Reickmann[2] say, there is no doubt about the role that universities must play in relation to sustainable development, through education and the development of competencies that will contribute to a more sustainable future.

Education for Sustainable Development (ESD) is a learning process based on the principles underlying sustainability. This process encompasses a broad range of experiences and programs. As stated in the Sustainability Education Handbook, published by UNESCO for the United Nations Decade of Education for Sustainable Development (The Manual of Education for Sustainability[22]), ESD is based on five types of learning to facilitate quality education and encourage Sustainable human development. Which are: "learn to know", "learn to be", "learn to live", "learn to do" and "learn to transform oneself and society". Education for Sustainable Development, in its broadest sense, is therefore linked to balanced development, in which it takes into account the social, cultural, environmental and economic dimensions of an improved quality of life for present and future generations.

The ESD is of particular importance in engineering education, which must include social, environmental and economic aspects to training socially responsible engineers, as they are basic in development society. Consequently, engineering education programs should be geared towards sustainability by bearing a part in the competencies between different disciplines, promoting interdisciplinary and sharing values among them to form for a sustainable future.

For this project, we have prepared and developed educational materials that allow working from Linear Algebra, issues related to sustainable development and social commitment. The material must be the basis for the preparation of projects that students must solve to assess this competency to all its dimensions. The competence about sustainable development and social commitment is understood as the ability to know and understand the complexity of social and economic phenomena that are typical of the welfare state: capacity to relate well with globalization and sustainability; capacity to use a stable and compatible technology, economy and sustainability (Guides to develop generic skills in designing degrees).

It will work the first level of achievement that is: systematic analysis and critical global situation, taking into account the sustainability of interdisciplinary and sustainable human development, and recognizing the social and environmental implications of the professional activity of the same field.

The other two levels of achievement that can be considered for undergraduate curricula consist of the following.

For level two: applying sustainability criteria and professional codes of conduct, in the design and assessment of technological solutions,

For level three: considering social, economic and environmental factors in the application of solutions, undertaking projects that tie in with human development and sustainability.

2. Acquisition of Sustainability Competence through the Subject of Linear Algebra

In engineering studies, the subject of Linear Algebra has been chosen as an essential and indispensable tool for anyone who deals with Mathematics, Arithmetic, Functional Analysis, Differential Geometry, Algebraic Topology, among others. In other fields, such as Engineering, Linear Algebra is used for example for the calculation of structures in which the notion of the eigenvalue is relevant. So it is a subject always thought to be applied in the longer term, making it difficult to think of introducing competencies other than those of the subject itself.

However, convinced of the importance of sustainability issues to be introduced from the outset, we have sought ways to make this possible.

Although, many people are convinced that Sustainable Development is completely disconnected with the mathematical reasoning. Nevertheless, they are very connected, and they need each other, and it is essential not only to show this relationship to the students but must work both competencies simultaneously. We will try to relate first-year college level math with one of the goals of sustainable development, concretely we chose the number six goal about clean water and sanitation.

Different mathematical models can be developed, for each one of the Sustainable Development goals. As, for diverse water problems such as the flow of rivers, lakes, reservoirs, drainage networks, supply networks, treatment plants, among others, for various renewable energy issues such as control of wind speed and direction as a means to estimate the region's wind power potential. These mathematical models are developed for better understand cause-effect relationships, to evaluate scenarios and to find alternative solutions to different problems related to the considered topic.

Many of these models are linear or linearized, so they can be treated using Linear Algebra.

2.1 Project Based Learning

Project-based learning (PBL) is a model that organizes the students learning around projects.

According to the definitions found in the PBL manuals for teachers, projects are complex tasks, based on ques-

 DOI: https://doi.org/10.30564/jiep.v1i1.317

tions or challenging problems, which involve students in the design, problem-solving, decision-making, or research activities. They give students the opportunity to work relatively autonomously for extended periods of time, and culminate in realistic outputs or result presentations (see (Jones et al.[11]; Thomas et al.[17], for example).

In a course where learning is based on projects, students apprehend about a topic by working for an elongated period to investigate and respond to a complicated question, challenge, or problem. PBL is an active method of learning which starts by posing questions, problems or scenarios rather than paper-based, rote memorization, or teacher-led instruction that presents facts (see Taberna and Garcia-Planas[16].

This method of learning presents at least the following advantages

Exploit student's autonomy

It allows to work collaboratively.

They come from the interests of the students and are motivators therefore.

They involve the development of core competencies.

Allow the work of different intelligences.

Allow the use of different learning strategies.

But, not all are advantages; we can also find some disadvantages.

They demand a paradigm shift in the way teachers think: we educate how they have educated us.

The concept of educational "level" needs to be redefined. Does the amount of knowledge or the quality of that knowledge matter?

Probably, if the change affects a whole matter or a center, it is necessary to convince the families of the outcome of the project.

They may need more teaching time to achieve the same academic goals.

It may at first confuse students.

Students who are bright or looking for more grades may be uncomfortable in a cooperative work group.

The biggest advantage is that in addition to the skills of collaboration, communication, critical thinking and the use of new technologies, it is possible to include work skills on sustainability. As a result, project-based learning helps students make their learning meaningful and rewarding by connecting them to the real world outside the classroom and generating confidence in the realization of their possible actions.

Then, it is clear that the use of PBL increases the significance of the learning that is, the relevance and the usefulness of what is learned.

For the PBL to be successful, the tasks to be carried out must be explained in detail, giving guides, advice, materials and offer opportunities for students to put them into practice through the tasks and activities proposed in the project.

On the other hand, students should be helped to overcome their tendency to postpone the process of completing the work by properly structuring the course.

2.1.1 How Does Project-Based Learning Work?

Project-based learning, as well as any learning method, requires a lot of planning. After a brainstorming, it begins with the selected idea and the approach of an essential question. Then many content standards will be addressed and devise a plan that will integrate as many subjects as possible into the project (see George Lucas Educational Foundation[10]).

Which can be summarized in the following steps:

- Start with the Essential Question.

The proposed question must be one that will engage the students.

- Design a Plan for the Project.

It is essential to have in mind which content standards will be addressed.

- Create a Schedule.

Design a timeline for project components taking into account:

- What time allotment will be given to the project?

- Will this project be conducted during the entire school day or during dedicated blocks of time?

- How many days will be devoted to the project?

Not forgetting that possible changes to the schedule will happen.

- Monitor the Students and the Progress of the Project.

Maintain control without preventing students from taking responsibility for their work, following these steps:

- Facilitate the process and the pleasure of learning.

- Encourage students to work collaboratively.

- Provide resources and guidance.

- Assess the process by creating team and project rubrics.

- Assess the Outcome.

Evaluate progress and give students feedback on how well they understand the information and on what they need to improve.

- Evaluate the Experience.

Reflection is a key component of learning. Allow for individual reflection, as well as group reflection and discussion.

The proposed question must be one that will engage the students.

Designing the project, it is essential that you have in mind which content standards will be addressed.

DOI: https://doi.org/10.30564/jiep.v1i1.317

2.2 Evaluation by Rubrics

Whenever there is an instruction, it is necessary to evaluate the acquired skills to know if the learning objectives have been achieved.

Evaluate is a process by which one or more characteristics of a student or a group of students are given the attention of the one who analyzes and evaluates their characteristics and conditions according to some criteria or reference points to emit a judgment that is relevant to the student's education.

Any method of evaluation employed must have clear and detailed criteria, which can be translated into parameters from which the evaluators will quantify and provide their assessment.

The methodology that we chose for the evaluation of the project is by rubrics because of it has been shown that the use of rubrics for assessment contributes significantly to improving students' performance.

The rubrics or rating matrices are "scoring guides used in student performance assessment that describe the specific characteristics of a product, project or task at various levels of performance, to clarify what is expected of the work of the Student, to evaluate its execution and to facilitate the proportion of feedback"

In the elaboration of a rubric, it is necessary to include the criteria of evaluation that gather the fundamental elements of the competence; they have to be criteria that offer to the student, with clarity, all the characteristics of the competence to develop[15]. More concretely, at the time of elaboration, it is convenient to take into account some aspects such as

Consider the characteristics of the competencies that have to be evaluated.

Determine accomplishment indicators (Development, originality neatness, creativity, clarity of writing, ability to synthesize, etc.).

Place the acquisition levels on a graduation scale to specify the differences in learning.

Determine the score awarded at each level.

The rubric must be available to students from the moment they call the project to be done. In this way, the students see that there is a system with some objectivity for the evaluation of their work and at the same time can evaluate their progress by consulting this rubric.

2.3 Tools for the Implementation of the Methodology

The progress made by information and communication technologies in recent years has had a significant effect on education and tools such as e-portfolio and others useful for e-learning have appeared; This tools can help improve and complement teaching Traditional in the classroom in many fields, including mathematics. In addition, they make it possible to make flexible the programs of the assignments facilitating the implementation of transversal competences.

In this case, we propose the use of the e-portfolio tool to support problem-based learning (PBL) in the teaching of the subject of Linear Algebra in which the transversal competence of sustainability and social commitment has been implemented.

An e-portfolio is a valuable tool for teaching, learning and assessment. Bear in mind that an e-portfolio is a digital collection of statement joined and managed by a user, usually on the Web. Such electronic evidence includes among others entering text, electronic files, images, multimedia, blog entries and hyperlinks. E-portfolios are a process of showing both the user's abilities and platforms for self-expression, and, if they are online, they can be maintained dynamically over time, (Domínguez-García et al.[7]).

Can be found various online platforms that allow creating a personal e-portfolio in general and educational e-portfolio in particular. Several authors (see (Bri et al.[3]; Domínguez-García et al.[8]) for example), analyse some of these educational platforms to facilitate the decision about which platform will be chosen.

One of the most used platforms is the open source e-Portfolio and social networking web application "Mahara" usually used joining an open source e-learning platform as for example Moodle. Both systems have built-in support for each other in the form of single sign-on and transfer of content or export different types of objects from Moodle to Mahara (API Portfolio) and to import objects from Mahara to Moodle (API repository)

Another open source of e-Portfolio able to be connected to Moodle by means of a plug is "Exabis". This platform permits us to introduce different pictures called categories where each of them gives access to different competencies. It is possible to access to multiple frames and assign different categories.

The Exabis platform is simpler than Mahara however is easier to import and export materials from Moodle.

As is well known, "Google Sites" is a free online application offered by the company Google. This application allows us to create a website in a simple editing a document. Both teachers and students can put in one place texts and images, as well as to include multiple documents. It also allows easy sharing personalized information across the network.

However, one limitation of this application is that the storage space is 100 MB. If the user is member if Google

 DOI: https://doi.org/10.30564/jiep.v1i1.317

Apps it can be expanded to 10 GB.

The tools for working on this platform provide graphic images in a more visually attractive way than Mahara and Exabis, making it more suitable for subjects such as drawing.

However, lately, it is spreading the use of the WordPress platform to create the portfolio for its sensibility in handling.

The use of learning platforms is an excellent tool for a quality education that is the basis for sustainable development, favoring access to education at all levels.

3. Using Case Study Methodology to Approach the Sustainable Water Management Competence

The Agenda 2030 for Sustainable Development (http://www.un.org/ sustainable development/en/), with an ambitious vision, includes seventeen Objectives integrating economic, social and environmental dimensions. Among these, there is the Goal 6: "Ensure access to water and sanitation for all". To achieve its aims is essential training to increase the efficient use of water resources in all sectors and ensure the sustainability of extraction and fresh water supply to cope with water scarcity and reduce the number of people suffering from water shortages.

To approach Sustainable Development Objectives in a mathematics subject, we make a teaching proposal based on projects.

Collaborative learning techniques are incorporated so that the project (that constitutes the core of the matter), be developed in the framework of project-based learning (PBL).

Despite the fact that the project has been implemented in large groups, it would seem natural to have opted for traditional master classes. This kind of courses offers certain advantages over other teaching methods because it is a fast, cheap and efficient method of transmitting information to a large number of students simultaneously. But it is also well known that student attention can only be maintained for short periods of 15 to 20 minutes separated by small spaces of 1-2 minutes in which the students stop paying attention; These periods of concentration are reduced, lasting less than five minutes at the end of a class, (Khan 2012).[12] In spite of everything, the master class has its meaning and can be used, but we think that it should not be the only method used in the classroom. So, we combine master classes with project-based teaching that helps to contribute effectively to the development of both cross-curricular and specific competencies that the subject must face.

The methodology of project-based learning enables different competencies to be worked in a cooperative way and can influence the understanding that a win is not sustainable, leading to the failure of others, which requires replacing competitiveness with cooperation. At the same time, it allows to contemplate environmental and developmental problems as a whole, taking into account their close relationship and the use of Linear Algebra, we can analyse their repercussions in the short, medium and long term.

Different mathematical models can be developed for various water problems such as a flow of rivers, lakes, reservoirs, drainage networks, supply networks, treatment plants, among others. These models are developed to understand cause - effect relationships better, to evaluate scenarios and to find alternative solutions to different problems related to water.

3.1 Project Proposal

This project is designed for a context of engineering school or faculty of science where a course of linear algebra is taught.

The proposal responds to the need to introduce the subject a competence that adds criteria and values consistent with sustainability and responds to the economic, social, cultural and environmental aspects of human development. Concretely, the subject considered is the water because it is a critical resource that has not substituted. Having potable water is a universal human right and is also a key factor for public health. The way we maintain and expand this critical good is a fundamental problem for building an environmentally and socially sustainable world.

This project can be applied in any studies in which contain the subject of linear algebra in its curriculum.

The following project is presented to students:

In a certain country, it is proposed to build a reservoir to regulate the basin of one of its rivers with the objective of satisfying the needs of water for irrigation.

For the realization of the project, the students are given the following data:

Maximum reservoir capacity,

Quantities required for irrigation

Volume to be left to maintain water quality standards for other uses, provided that the water level of the dam plus the weekly contribution by the water of the river, does not reach a minimum that does not allow the exit of water.

The primary objective to be achieved with this project is: "Study the viability of the reservoir" by analyzing

The stability of the reservoir under the given conditions

The sustainability of the reservoir under the same conditions

DOI: https://doi.org/10.30564/jiep.v1i1.317

The stability of the reservoir imposing the variation in the time of the river's contribution to climate change.

The sustainability of the reservoir by imposing the previous condition

It is also intended that the following awareness-raising objectives be achieved:

Make an assessment of the social benefit by counterbalancing the benefits to be obtained by irrigation in the face of social conflict caused by expropriations and the resulting displacement of the inhabitants of the area.

Make an assessment of the problem by extrapolating the case to a large dam.

To solve the project, the student must:

Describe using a matrix equation of the type p(k + 1) = Ap(k), the weekly transition of the probable water units.

2. Express p(k) as a function of p(0)

Starting from a particular amount of water impounded:

Analyze the probability that in two weeks the reservoir will be below minimums.

Critical assessment of the result within the context of the work.

Find out how Linear Algebra will give us a solution to the matrix equation proposed

Apply it to the case at hand

Use the calculations to analyze the situation of the reservoir at the week

k = 10.

Study and analysis of stability and sustainability.

The organized task for PBL is precisely aligned with one of the Sustainable Development Objectives 2030, more concretely with the goal number six "Ensure access to water and sanitation for all". To achieve this aim is essential training to increase efficiency of water resources in all sectors and ensure the sustainability of freshwater harvesting and water supply to address water scarcity by reducing the number of people suffering from water shortages.

With this project not only introduces and evaluates the competence on sustainability but also the social commitment because they must also value the social cost that involves the displacement of people with the uprooting that this entails.

Clearly, the students need to learn a full linear algebra course to solve completely and in a correct form, the project.

3.1.1 Course Planning

Suppose that the subject of linear algebra corresponds to a course of 150 hours of which 40% corresponds to the work done in the classroom and the remaining 60% to work done outside the school. Theoretical classes occupy 25% of the time devoted to the classroom tasks, and complete the time with tutorials and solving the doubts that have arisen for the students. In addition, students work collaboratively on the proposed project related to the academic content of the subject. This accounting has been realized taking into account the European Credit Transfer system, (ECTS). In this system are also assessed the hours that students devote to the self-activity. The computation corresponds to 25 hours per credit, and the subject is valued at six credits. One of the significant advantages of the ECTS credits is that being the unit of measure equal in all the universities of the new European space allows comparing much more efficiently the educational load of the degrees.

We plan, as can be seen in Figure 1, the distinct tasks that must be solved by the students through the different weeks of the course to achieve with success the project.

Figure 1. Subject Planning

The plan has been prepared conscientiously having in mind the progress of the lectures in which they are introduced the theoretical support corresponding to the linear algebra subject and that should serve them as the basis for the completion of the project.

Students are given the work plan at the same time that they are notified of the project they are required to complete.

The students are invited to follow the plan as closely as possible so that they can carry out their project progressively and profitably.

3.1.2 Evaluation

When designing a subject one of the fundamental questions that one has to think about is how what and when to evaluate. If we focus the evaluation on the competencies that we want the student to acquire, we can give coherence to the structure, approach and the development of the subject. Keep in mind that they are first-year college students and the first project they face, so it is important to help them plan to avoid failure.

Following Monereo[13], "Say you how you evaluate, and I will tell you how your students learn" the evaluation has a retroactive role on learning and teaching because it modifies the way of learning and educating.

Once we have defined the competencies that we want to evaluate and have designed the methodology to front

 DOI: https://doi.org/10.30564/jiep.v1i1.317

learning for the acquisition of these skills we have to proceed with the evaluation.

As we have indicated in subsection 3.2, the evaluation of the project is done by rubrics.

We have valued the current rubrics, and we have found that they do not provide the required information, for Linear Algebra course, in general, and for mathematics topic including evaluation of transversal competencies as sustainability. So we have chosen to perform new rubrics in such a way that they perfectly adapted to the course, which includes all skills treated, that is evaluated. Table 1 Presents the Part of Rubric Corresponding to the Evaluation of Sustaina- bility.

Table 1. Evaluation of the transversal competency involved in the PBL

Task	descriptors			
Description of the water sustainability problem	NO correct identification 1 point	YES, identifies the problem, but NO properly 2 points	YES, identifies the problem, but something confusing 3 points	YES, identifies the problem, but clearly and accurately 5 points
Analysis of stability and sustainability	Has not been able to perform the study 1 point	Has not been able to analyze stability and sustainability 2 points	Has correctly analyzed only one of the concepts 4 points	Has analyzed the problem correctly 7 points

Sustainability competence is evaluated in each of the evaluation acts, in the examinations appear detailed questions linked to the competencies of sustainability.

The project values the treatment of sustainability in a special way, although in each and every one of the points to be evaluated, this concept is implicit. Also, different aspects of the project are assessed, the ones related to the formal appearance and care in the presentation of the work, the solving process of the project and the self-assessment and peer assessment of the student work. Self-evaluation is a good strategy to increase responsibility and for students to learn to value, criticize and reflect on their learning process and peer evaluation enables students to evaluate peer work and compare work that leads them to improve their realized work.

Due to the good results, we have prepared text with different projects to continue to implement in class (Garcia-Planas et al. [9]).

4. Results

This project has been implemented by the students of the Barcelona School of Industrial Engineering (ETSEIB) from the Polytechnical University of Catalonia (UPC), on the subject of linear algebra during the academic year 2016-17. The results in the process of acquisition of the transversal competence on sustainability and social commitment on the part of the students are satisfactory. Specifically regarding the objective of sustainable development "Sustainable water management" students through project-based learning implemented in the subject of Linear Algebra, have become aware and have been able to address, deepen and disseminate the proposed theme in the classroom through the virtual campus of the UPC.
It is important to emphasize the change of attitude of students with regard to the ones of the previous years.

5. Conclusion

Sustainability is often emphasized as an essential objective of higher education, but more as a principle than at a practical level. This work has shown how it is possible to carry out the implementation of sustainability competence in an evaluable form, in a subject of higher education mathematics.

As a first conclusion, we have that it is imperative that universities in making their curricula take into account sustainable development. Little by little, the high schools are becoming aware but the process is slow and according to which subjects the teacher is less open to including this competence in his subject.

With this work, we have proved that it is possible to implement the competition "Sustainability and social commitment" within a mathematics subject, to address the objectives of the Agenda 2030 on sustainable development.

For this, it has been important to see how cross-sectional looking for real problems, which the student must model, it is possible to evaluate such competence.

In this case the analysis of the sustainability of water management. It is important to emphasize the change of attitude of the students about previous years.

Collaborative learning techniques are incorporated to ensure that the project that constitutes the core of the subject is developed in the framework of project-based learning (PBL).

The presented project has significance since the learning through PBL implies changes in the knowledge structures of the students, modifies them and enriches them since it establishes new connections and relations between them. In the learning process, the student develops a significant relationship between what has been acquired and everything that must be learned and assimilated through the project "connecting" the new information with relevant concepts previously acquired.

The project has contributed to the development of an organizational culture that fosters principles such as meaningful learning, collaboration, responsibility, innovation, mutual help and respect, participation, individual and collective reflection.

References

[1] Allen, D., Allenby, B., Bridges, M., Crittenden, J., Davidson, C., Matthews, S., Murphy, C., Pijawka, D. and Nolt-

  DOI: https://doi.org/10.30564/jiep.v1i1.317

helm, C. (2008). Benchmarking Sustainability Engineering Education: Final Report: EPA Grant X3-83235101-0.

[2] Barth, M., Rieckmann, M. (2012). Academic staff development as a catalyst for curriculum change towards education for sustainable development: an output perspective, Journal of Cleaner Production 26, 28-36.

[3] Bri, D., García, M., Coll, H. LLoret, J. 2009. A Study of Virtual Learning Environments. Wseas Transactions on Advances in Engineering Education, 6 (1), 33-43.

[4] Colombo, C.R., Alves, A.C., van Hattum-Janssen, N., Moreira, F. (2014). Active Learning Based Sustainability Education: a Case Study. Proceedings of the Sixth International Symposium on Project Approaches (PAEE2014) Medellin Colombia.

[5] Colombo, C.R., Moreira, F., Alves, A.C. (2015). Sustainability Education in PBL Education: the case study of IEM-UMINHO. CGIT - Publicaçes em actas de encontros científicos/Papers Conference proceedings 221-228.

[6] Crofton, F.S. (2000). Educating for sustainability: opportunities in undergraduate engineering. Journal of Cleaner Production, 8, 397-405.

[7] Domínguez-García, S., García-Planas, M.I., Taberna, J. (2016). Mathematical modelling in engineering: an alternative way to teach Linear Algebra. International Journal of Mathematical Education in Science and Technology, 47 (7), 1076-1086.

[8] Domínguez, S., Garcia-Planas, M.I., Palau, R., Taberna, J. (2015). Modelling E-portfolio for a Linear Algebra undergraduate course. International journal of education and information technologies, 9, 115-121.

[9] García-Planas, M.I., Taberna, J., Rina, N. (2018). Linear algebra in education for sustainable development [Álgebra lineal en la educación para el desarrollo sostenible]. Barcelona, Spain: Iniciativa Digital Politècnica.

[10] George Lucas Educational Foundation. 2007 Edutopia. How Does Project-Based Learning Work? Tools for understanding the process of planning and building projects. https://www.edutopia.org/project-based-learning-guide-implementation (Retrieved 21/11/2017)

[11] Jones, B.F., Rasmussen, C. M., Moffitt, M. C. (1997). Real-life problem solving.: A collaborative approach to interdisciplinary learning. Washington, DC: American Psychological Association.

[12] Khan, S. (2012). Why Long Lectures Are Ineffective. Time Ideas. Retrieved from http://ideas.time.com/2012/10/02/why-lectures-are-ineffective/

[13] Monereo, C. 2009. Evaluation as a tool for educational change [La evaluación como herramienta de cambio educativo]. In PISA as an excuse. Rethink the evaluation to change the teaching [En PISA como excusa. Repensar la evaluación para cambiar la enseñanza]. Barcelona: Graó, pp. 15--31.

[14] Murphy, C.F., Allen, D., Allenby, B., Crittenden. J., Davidson, C., Hendrickson, C., Matthews, H.S. (2009). Sustainability in Engineering Education and Research at US Universities, Environ. Sci. Technol., 43 (15), 5558--5564.

[15] Taberna, J., Domínguez Garcia, S., García-Planas, M.I. 2016. The use of tic for an evaluation in competences in higher education [El uso del tic para una evaluaci\'on en competencias en la educaci\'on superior]. XIII Fecies

[16] Taberna, J. Garcia-Planas, M.I. 2016. Project-based teaching: an alternative way to teach Linear Algebra [Enseñanza basada en proyectos: una forma alternativa para enseñar Álgebra Lineal]. Proceedings IN-RED 2016 - II Congreso Nacional de Innovación Educativa y de Docencia en Red INRED UPV, 826-839.

[17] Thomas, J. W., Mergendoller, J. R., Michaelson, A. (1999). Project-based learning: A handbook for middle and high school teachers. Novato, CA: The Buck Institute for Education.

[18] Trimingham, R., Lazzarini, B., Pérez-Foguet, A., Noble, N., Boni, A., Sierra Castañer, M., Mongera, F., Zolezzi, G. (2016). Integrating the Global Dimension in Engineering Education: Experiences from a Collaborative Project. In Teaching Education for Sustainable Development at University Level, Book Chapter, Springer, Switzerland, pp. 175--190. doi: 10.1007/978-3-319-32928-4_12

[19] Wells, P.J. (2017). UNESCO Higher Education Priorities within the context of the Sustainability Goals. CHEA International Quality Group, Washington D.C. https://www.chea.org/userfiles/Conference%20Presentations/2017-ciqg wells. pdf

[20] Comisión Brundtland (1987) Report of the World Commission on Environment and Development Our Common Future.

[21] Guides to develop generic skills in designing degrees. Retrieved May 22, 2017 from https://www.upc.edu/ice/ca/innovacio-docent/publicacions_ice.

[22] The Manual of Education for Sustainability. United Nations Decade of Education for Sustainable Development (2005-2014). Retrieved january 20, 2017 from http://www.urv.cat/media/upload/arxius/catedradesenvolupament sosteni ble/Informes%20VIP/unescetxea-manual_unesco_cast_education_for_sustaina bility_manual.pdf

[23] Sustainable development goals. 17 goals to transform our world. Retrieved on January 15, 2017 from http://www.un.org/ sustainabledevelopment/es/

 DOI: https://doi.org/10.30564/jiep.v1i1.317

Author Guidelines

This document provides some guidelines to authors for submission in order to work towards a seamless submission process. While complete adherence to the following guidelines is not enforced, authors should note that following through with the guidelines will be helpful in expediting the copyediting and proofreading processes, and allow for improved readability during the review process.

Ⅰ. Format

- Program: Microsoft Word (preferred)
- Font: Times New Roman
- Size: 12
- Style: Normal
- Paragraph: Justified
- Required Documents

Ⅱ. Cover Letter

All articles should include a cover letter as a separate document.

The cover letter should include:

- Names and affiliation of author(s)

The corresponding author should be identified.

Eg. Department, University, Province/City/State, Postal Code, Country

- A brief description of the novelty and importance of the findings detailed in the paper

Declaration

v Conflict of Interest

Examples of conflicts of interest include (but are not limited to):

- Research grants
- Honoria
- Employment or consultation
- Project sponsors
- Author’s position on advisory boards or board of directors/management relationships
- Multiple affiliation
- Other financial relationships/support
- Informed Consent

This section confirms that written consent was obtained from all participants prior to the study.

- Ethical Approval

Eg. The paper received the ethical approval of XXX Ethics Committee.

- Trial Registration

Eg. Name of Trial Registry: Trial Registration Number

- Contributorship

The role(s) that each author undertook should be reflected in this section. This section affirms that each credited author has had a significant contribution to the article.

1. Main Manuscript

2. Reference List

3. Supplementary Data/Information

Supplementary figures, small tables, text etc.

As supplementary data/information is not copyedited/proofread, kindly ensure that the section is free from errors, and is presented clearly.

Ⅲ. Abstract

A general introduction to the research topic of the paper should be provided, along with a brief summary of its main results and implications. Kindly ensure the abstract is self-contained and remains readable to a wider audience. The abstract should also be kept to a maximum of 200 words.

Authors should also include 5-8 keywords after the abstract, separated by a semi-colon, avoiding the words already used in the title of the article.

Abstract and keywords should be reflected as font size 14.

Ⅳ. Title

The title should not exceed 50 words. Authors are encouraged to keep their titles succinct and relevant.

Titles should be reflected as font size 26, and in bold type.

Ⅳ. Section Headings

Section headings, sub-headings, and sub-subheadings should be differentiated by font size.

Section Headings: Font size 22, bold type

Sub-Headings: Font size 16, bold type

Sub-Subheadings: Font size 14, bold type

Main Manuscript Outline

Ⅴ. Introduction

The introduction should highlight the significance of the research conducted, in particular, in relation to current state of research in the field. A clear research objective should be conveyed within a single sentence.

Ⅵ. Methodology/Methods

In this section, the methods used to obtain the results in the paper should be clearly elucidated. This allows readers to be able to replicate the study in the future. Authors should ensure that any references made to other research or experiments should be clearly cited.

Ⅶ. Results

In this section, the results of experiments conducted should be detailed. The results should not be discussed at length in

this section. Alternatively, Results and Discussion can also be combined to a single section.

Ⅷ. Discussion

In this section, the results of the experiments conducted can be discussed in detail. Authors should discuss the direct and indirect implications of their findings, and also discuss if the results obtain reflect the current state of research in the field. Applications for the research should be discussed in this section. Suggestions for future research can also be discussed in this section.

Ⅸ. Conclusion

This section offers closure for the paper. An effective conclusion will need to sum up the principal findings of the papers, and its implications for further research.

Ⅹ. References

References should be included as a separate page from the main manuscript. For parts of the manuscript that have referenced a particular source, a superscript (ie. [x]) should be included next to the referenced text.

[x] refers to the allocated number of the source under the Reference List (eg. [1], [2], [3])

In the References section, the corresponding source should be referenced as:

[x] Author(s). Article Title [Publication Type]. Journal Name, Vol. No., Issue No.: Page numbers. (DOI number)

Ⅺ. Glossary of Publication Type

J = Journal/Magazine

M = Monograph/Book

C = (Article) Collection

D = Dissertation/Thesis

P = Patent

S = Standards

N = Newspapers

R = Reports

Kindly note that the order of appearance of the referenced source should follow its order of appearance in the main manuscript.

Graphs, Figures, Tables, and Equations

Graphs, figures and tables should be labelled closely below it and aligned to the center. Each data presentation type should be labelled as Graph, Figure, or Table, and its sequence should be in running order, separate from each other.

Equations should be aligned to the left, and numbered with in running order with its number in parenthesis (aligned right).

Ⅻ. Others

Conflicts of interest, acknowledgements, and publication ethics should also be declared in the final version of the manuscript. Instructions have been provided as its counterpart under Cover Letter.

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice

Aims and Scope

Journal of International Education and Practice is devoted to the study and exploration of education, reading, curriculum research, education policy, sociology of education, and economics.

It covers all innovations of educational research, including but not limited to:

- Basic education theories
- Education policies and regulations
- Education management
- Curriculum and teaching
- Examination and evaluation
- Education economy
- Education psychology
- Higher education
- International and comparative education
- Academic essays
- Education technology

Bilingual Publishing Co. (BPC)

Tel:+65 65881289

E-mail: contact@bilpublishing.com

Website: www.bilpublishing.com

About the Publisher

Bilingual Publishing Co. (BPC) is an international publisher of online, open access and scholarly peer-reviewed journals covering a wide range of academic disciplines including science, technology, medicine, engineering,education and social science. Reflecting the latest research from a broad sweep of subjects, our content is accessible worldwide – both in print and online.

BPC aims to provide an analytics as well as platform for information exchange and discussion that help organizations and professionals in advancing society for the betterment of mankind. BPC hopes to be indexed by well-known databases in order to expand its reach to the science community, and eventually grow to be a reputable publisher recognized by scholars and researchers around the world.

BPC adopts the Open Journal Systems, see on http://ojs.bilpublishing.com

Database Inclusion

Asia & Pacific Science Citation Index

Creative Commons

China National Knowledge Infrastructure

Google Scholar

Crossref

MyScienceWork

Bilingual Publishing Co. (UEN: 32307000X)
has maintained a prominent influence
in Southeast Asia for over 30 years.
Our mission is to advance academic research of the world
by publishing the latest research and encouraging discourse
across various scientific disciplines.
The main office of Bilingual Publishing Co
is located at 12 Eu Tong Sen Street, #07-169, Singapore 059819.

BILINGUAL
PUBLISHING CO.
Pioneer of Global Academics Since 1984

Tel: +65 65881289
E-mail: contact@bilpublishing.com
Website: www.bilpublishing.com

ISSN 2630-516X
01
9 772630 516183