

Design and Validation of the College Readiness Test (CRT) for Filipino K to 12 Graduates

Antonio Tamayao¹, Rudolf Vecaldo¹, Jay Emmanuel Asuncion², Maria Mamba³, Febe Marl Paat⁴, Editha Pagulayan⁵

¹ College of Teacher Education, Cagayan State University, Andrews Campus, Tuguegarao City, Cagayan, Philippines

² College of Allied Health Sciences, Cagayan State University, Andrews Campus, Tuguegarao City, Cagayan, Philippines

³ College of Arts and Sciences, Cagayan State University, Carig Campus, Tuguegarao City, Cagayan, Philippines

⁴ Graduate School, Cagayan State University, Carig Campus, Tuguegarao City, Cagayan, Philippines

⁵ Graduate School, Cagayan State University, Andrews Campus, Tuguegarao City, Cagayan, Philippines

Correspondence: Rudolf Vecaldo, College of Teacher Education, Cagayan State University, Tuguegarao City, 3500, Cagayan, Philippines. E-mail: rudolfvecaldo@gmail.com

Received: December 16, 2019

Accepted: January 31, 2020

Online Published: February 4, 2020

doi:10.5430/ijhe.v9n2p209

URL: <https://doi.org/10.5430/ijhe.v9n2p209>

Abstract

Designing and validating a college readiness test addresses the absence of standardized Philippine-based College Readiness Test (CRT) congruent with the College Readiness Standards (CRS) set by the Philippine Commission on Higher Education (CHED). It also resolves the varied and arbitrary indices used by Higher Education Institutions (HEIs) to measure the preparedness of K to 12 Filipino graduates to enter college. In this regard, this study establishes the validity and reliability of the CRT to measure the combination of knowledge, skills, and reflective thinking necessary for the K to 12 graduates to be admitted and to succeed without remediation in the General Education courses in HEIs. Using multi-stage sampling in a select province of the Philippines and with due consideration of the district, type of school, and academic tracks offered in senior high school, the study has generated that the 200-item CRT has *desirable* difficulty index (65.64), *reasonably good* discrimination index (0.22), and large functioning distractors (68.91% distractor efficiency). Notably, there is a significant positive relationship between discrimination and difficulty indices as well as the distractor efficiency and difficulty index of the CRT items. Also, the CRT is reliable as it possesses inter-item consistency ($r=0.796$). Thus, it is a valid and reliable instrument to measure the college readiness of Filipino K to 12 graduates with its features of being contextualized, gender-fair, and criterion-referenced.

Keywords: college readiness, item analysis, validity, reliability, K to 12 graduates

1. Introduction

The changing landscape of education in the 21st century demands learners to acquire competencies necessary to adapt and to survive in the growing information age. In this respect, various countries have aimed their basic education systems to produce graduates equipped with the knowledge and skills in becoming globally competitive and lifelong learners, and to develop citizens who can significantly contribute to the emerging economic, socio-cultural and political affairs (Dougherty, Mellor, & Smith, 2006; Merisotis & Phipps 2000). In response to these demands, the Philippines has intensified its national development thrusts. It has embarked on curricular reforms to significantly enhance its educational standards and outcomes, alongside its efforts to address its pressing problems and concerns in the education system (Okabe, 2013). The comprehensive re-engineering of the Philippine basic education was undertaken by the government by adding two years of schooling from 10 to 12 years (ICEF, 2013). This situation prompted the creation of K to 12 Education Program, which was institutionalized through Republic Act No. 10533 (Enhanced Basic Education Act of 2013). This major reform intends to decongest the offering of subjects for students to acquire mastery of skills and competencies required in the 21st century, such as literacy, critical thinking, and technical skills. These skills are hoped to enable the senior high school graduates to be prepared for work or tertiary education (Magno & Piosang, 2016).

Concomitant to the changes in the basic education curriculum, the Philippine government also reinforced its serious

commitment to further implement its educational plan through reforms instituted in the higher education level. Hence, the Commission on Higher Education (CHED), being the regulatory agency for both private and public institutions of higher learning, has set mechanisms to ensure the congruence and synergy of learning goals between the basic education and higher education. With this, the Commission came up with the College Readiness Standards (CRS) as the overarching framework in ensuring quality and preparedness of postsecondary students to undertake tertiary education and in modifying the general education curriculum offered in college (CEB Resolution No. 298-2011). The Technical Panel on General Education (TPGE) defines CRS as follows:

College Readiness Standards (CRS) consists of the combination of knowledge, skills, and reflective thinking necessary to participate and succeed – without remediation – in entry-level undergraduate courses in Higher Education Institutions (HEIs).

The college readiness aims to determine who among the K to 12 graduates are fit to enter college (Jawad, 2017) since a high school diploma may not necessarily reflect adequate preparation “for the intellectual demands of adult life” (Porter & Polikoff, 2012: 401). In Conley’s (2007) framework, the components of academic preparedness are: (1) content knowledge; (2) academic skills i.e. a student’s application of content knowledge to solve problems; and (3) key cognitive strategies, such as reasoning, arguing, and interpreting. In essence, college readiness is the interplay of the content knowledge, practical skills and disposition to withstand the unique challenges of tertiary education (Conley, Aspengren, Stout, & Veach, 2006) making college readiness as a vital construct in determining academic success in a post-secondary education.

Among leading universities and colleges, a well-established and credible means of assessing college readiness are the standardized tests. These standardized tests have shown their predictive value to student’s college attendance, persistence to graduation, and academic performance. For example, students who passed the standardized test such as Scholastic Assessment Test (SAT) and American College Test (ACT) have shown likelihood to enroll in tertiary education (Avery and Kane, 2004). Their scores are highly predictive not only of college enrolment (Roderick, 2006), but also of academic achievement during first year in college (Burton & Ramist, 2001).

With the operationalization of the K to 12 program in the Philippines, the Department of Education (DepEd) has graduated its first batch of senior high school graduates in 2018. As part of the assessment mechanism, the Department administered the Basic Education Exit Assessment (BEEA) purposely to measure the competencies acquired by the students in Grades 11 and 12 only. However, it is not intended to ascertain the college readiness of the senior high school graduates. To date, there is no standardized Philippine-based College Readiness Test (CRT) to measure the preparedness and possible success of these graduates based on the CRS set by CHED. Consequently, admission of first year college students in higher education institutions solely depends on varied and arbitrary measures. This prevailing situation creates diverging views on the construct of college readiness which may not reflect the competencies acquired by senior high school graduates as defined by the CRS.

According to Porter & Polikoff (2012:402), “a well-designed and validated readiness assessment could help schools, districts, and states provide more uniform expectations for students across the country.” It is in this premise that this study was conducted to address the foregoing gap. Specifically, it aimed to: (1) design and construct a criterion-referenced, contextualized, gender-fair Filipino College Readiness Test (CRT) congruent with the CRS; (2) establish the difficulty index, discrimination index, and distractor efficiency of the CRT; (3) ascertain the association of CRT indices; and (4) determine the reliability (inter-item consistency) of the CRT.

2. Literature Review

2.1 Delineating the Constructs of College Readiness

The remarkable changes in the educational landscape call for greater attention in making secondary students being college ready after graduation (Conley, 2010). Hence, educational leaders remain focused in understanding how students can be adequately prepared with the knowledge, skills, and attitudes needed for them to be admitted and to succeed in college (Barnes, Slate & Rojas-LeBouef, 2010; Barnes & Slate, 2013). This situation leads to the determination of various conceptions on college readiness among researchers and policy-makers in order to establish mechanisms for a smooth transition from high school to tertiary education.

To date, researchers in the fields of educational psychology and education policy have not reached a common definition of college readiness (Osion, 2006). Several frames of understanding college readiness surface depending on the level of analysis being focused on. According to Porter & Polikoff (2012), college readiness may be understood as those relating to non-cognitive factors such as student personality traits, grit, and family resources. On the contrary, some researchers such as Bridgeman (1991) and Kobrin, Patterson, Shaw, Mattern, & Barbuti (2008)

attribute college readiness to academic performance, which comprises grade point average, class rank, and scores in various academic tests.

Along this line, Greene and Foster (2013:3) lay the basic definition of college readiness as “the bare minimum qualifications necessary before the college will even consider the student’s application”. In other words, college readiness highlights the admission of students in the tertiary level. This notion could be further reinforced using Conley’s (2007:5) framework wherein college readiness is “the level of preparation a student needs in order to enroll and succeed, without remediation, in a college bearing education course at post-secondary institution that offers a baccalaureate degree or transfer to a baccalaureate program.” With this position, college readiness is operationally viewed as the ability of the student not just to be accepted in tertiary level but to succeed in college life as well.

Going beyond college eligibility as indicator of college readiness, Kless, Soland, & Santiago (2013) argued that college readiness may be construed in three levels, namely, individual, setting and systems. Individual level focuses on the typical markers of college-bound student such as but not limited to study skills, courses and credits taken, persistence, and expectations for the future. Setting level, on the other hand, refers to the opportunities and resources provided to the students by the academic institution. Lastly, systems level highlights “policy and funding infrastructure that impacts school resources, student programs, and college readiness supports” (Kless, Soland, & Santiago, 2013:5).

2.2 Various Measures of College Readiness

College readiness is an emerging educational construct and in most countries, it has become a major national challenge (Leonard, 2013). As a construct, it has been explored using multiple measures. Several studies claim that the grade point average (GPA) in high school is a measure of readiness (Camara, & Echternacht, 2000; Reuschel, 2009). Other related indicator in western countries is the rigorous work load wherein students who took more academic coursework, such as four years of English; three years each of natural science, social science, and math; and two years of foreign language were promoted to higher years (Greene & Foster, 2003) and successfully completed a college degree with 20 credits of course work in the first year of college (Adelman, 2006).

On the other hand, other researchers have proposed the use of questionnaire or checklist as a way of gauging one’s readiness to enter college. For instance, in Landmark College, a questionnaire is utilized to measure college readiness by determining students’ academic skills, self-understanding, self-advocacy, execution function, and motivation. This tool ascertained the set of skills, practices, and competencies of the foregoing dimensions which are associated to the preparedness of the students with disabilities in their entry to tertiary education (Landmark College, 2009).

2.3 College Readiness Standards in the Philippines

In 2011, CHED came up with College Readiness Standards (CRT) in order to lay the competencies that are expected to be developed and mastered by the graduates of the K to 12 Program as essential entry competencies in college. These standards were prudently drafted by the Technical Working Group composed of experts from DepEd, CHED, and representatives from private and public higher education institutions (CHED En Banc Resolution No. 298-2011).

There are seven (7) learning areas in the CRS, namely, English, Filipino, Literature, Mathematics, Science, Social Studies, and Humanities. Each learning area is composed of both content and performance standards. Content standards refer to what the students are expected to know whereas the performance standards refer to what the students are expected to demonstrate with what they know. As stipulated in CHED En Banc Resolution No. 298-2011 the CRS intends to “remove the remedial character of entry-level college courses and to allow higher learning institutions to tighten the focus of their undergraduate curricula, as well as to conform to international standards.”

To date, there is no assessment done as regards college readiness vis-à-vis the CRS framework. In the basic education sector, an exit test was administered to Grade 12 graduating students known as the Basic Education Exit Assessment (Mateo, 2019). This test focuses only on the competencies of senior high school, thus it is not comprehensive and may not be constructively aligned with the CRS. Meanwhile, in order to address the issue on college readiness, CHED laid CMO No. 10, Series of 2017 as guide for higher education institutions in the Philippines in admitting their first year students. In this memorandum, a bridging program is expected to be implemented by universities and colleges in addition to their institutional admission policies and requirements (Taghoy, 2019).

2.4 Conceptual Framework

The existing Philippine CRS is the articulation of competencies that build on the learning outcomes of the basic education and the demands of international standards. In order to assess the degree to which Filipino K to 12 graduates have mastered the content necessary to be college ready, a well-designed and validated College Readiness Test is indispensable. This concern is imperative considering that a standardized test is one of the credible means of assessing the college readiness of K to 12 graduates. However, any standardized test must establish its validity and reliability prior to its use. Hence, the CRT, a criterion-referenced test, was designed and validated measuring the college readiness of K to 12 graduates along their achievement in the seven (7) learning areas consisting of the content and performance standards defined in the CRS. Given such tool, it shall measure their knowledge, skills, and reflective thinking necessary to be accepted in college and to succeed in foundational courses. This is with the end in view of providing meaningful feedback for the enhancement of college readiness in the Philippines which is vital in evaluating the educational improvement and achievement gaps of K to 12 graduates overtime. Figure 1 shows the conceptual paradigm of the study.

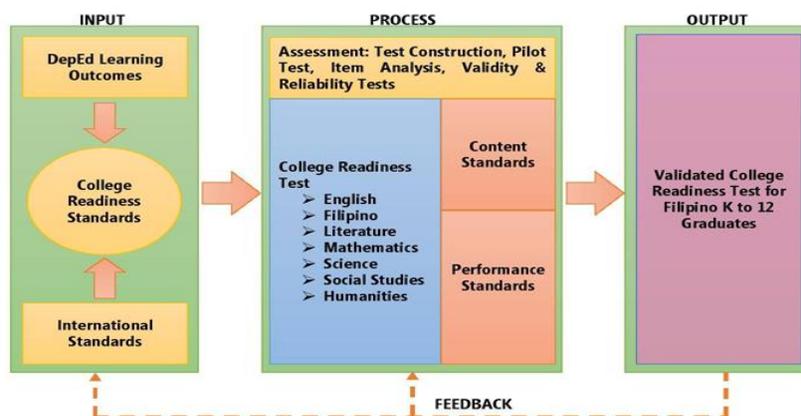


Figure 1. Conceptual Paradigm of the Study

3. Methods

3.1 Research Design

Quantitative design was used in the study specifically ascertaining the validity and reliability of the College Readiness Test (CRT). Validity is integral in ensuring the instrument's trustworthiness and accuracy in measuring the intended outcomes (Bond, 2003). It establishes the empirical evidences and theoretical rationales of an instrument as it measures what it purports to measure. This leads to adequacy and correctness of interpretations based on test scores which strengthens the claims and inferences made by researchers about the results of an assessment (Messick, 1989). On one hand, reliability reflects the consistency and stability of the instrument over series of tries. An instrument is reliable when it is applied to the same object of measurement several times but yields similar results (Fraenkel & Wallen, 2003; McMillan & Schumacher, 2001; Moss, 1994; Neuman, 2003). These two constructs were vital in establishing the credibility of the CRT as a criterion-referenced test.

3.2 Sampling Procedure

Multistage sampling technique was employed in the first and second pilot tests to obtain the target participants. The first stage of sampling considered the three districts of the select province of the Philippines by randomly selecting the towns offering senior high school in each district. The second stage of sampling considered the type of school, (i.e. public and private) for realistic representation of study participants. The final stage of sampling randomly selected the graduating senior high school students using the school register in due consideration of all the offered senior high school tracks in the province.

3.3 Research Participants

3.3.1 First Pilot Test

The CRT was first pilot-tested to 231 graduating students of the senior high school in one Philippine province for the School Year 2018-2019. These students came from five (5) public (big-sized and small-sized) and three (3) private (sectarian and non-sectarian) senior high schools. Big-sized schools are those that offer more tracks with more resources and bigger teacher-student population whereas small-sized schools are those that offer only one track with less resources and smaller teacher-student population. Using the school register as a sampling frame, the participants were selected through stratified random sampling considering all tracks in Senior High School namely, academic (116 participants from STEM, HUMMS, ABM, and GAS), technical-vocational (81 participants from agri-fishery, home economics, industrial arts and ICT), and sports (34 participants from safety and first aid, fitness testing, fundamentals of coaching and human movement).

3.3.2 Second Pilot Test

The second pilot test had 942 participants drawn from 10 public schools and 8 private schools located in the three (3) congressional districts of the select province. These participants were chosen through stratified random sampling in consideration of the following tracks: academic (565 participants from STEM, HUMMS, ABM, and GAS), technical-vocational (377 participants from agri-fishery, home economics, industrial arts and ICT), and sports (94 participants from safety and first aid, fitness testing, fundamentals of coaching and human movement).

3.4 Research Procedure

The CRT was developed through the following phases:

1. **Writing Phase** – The test items were constructed by select experts teaching in the senior high schools of the Department of Education (DepEd) and general education courses in the public higher education. This strategy ensured that the test items cut across the competencies taught in the basic education and those found in the College Readiness Standards (CRS). Test constructors were further capacitated through a writeshop for familiarization with the competencies reflected in the CRS and the objectives for developing the CRT. Seven (7) learning areas, namely, English, Filipino, Literature, Mathematics, Science, Social Studies, and Humanities were considered in developing the CRT as defined in the CRS. The Table of Specifications (TOS) per learning area was developed by mapping out the essential entry competencies (requisites) to hurdle the general education courses in the tertiary level. Thirty-percent (30%) of the test items were content standards (knowledge and comprehension) and 70% were performance standards i.e., application, analysis, evaluation, testing assumptions, hypotheses, and relevant information. Each of the seven (7) learning areas has fifty (50) items consisting of four (4) options totaling to 350 items. A template guided each test constructor in determining the tested competency, the cognitive level as well as the perceived item difficulty, plausibility of distractors, and key to correction. Finally, contextualization, and gender-fair language were considered in constructing the test items to make them more meaningful and significant to the lives of the test takers.
2. **Content Validation** – The output of the writing phase was subjected for content validation by subject experts from a select public higher education institution offering comprehensive programs capturing all the academic tracks offered in basic education. This ensured that the test items were aligned with the competencies reflected in the CRS and they measure the competencies tested in the CRT. Content validators were provided with a rubric in validating the test items. Out of 350 items, 50 were rejected because they were considered poor. Moreover, the comments and suggestions of the content validators were used in improving the CRT for the first pilot test. Improvements in the items were either simplification of the stem and stimulus material or improvement of the options or distractors.
3. **First Pilot Test** – The first pilot test was a dry-run as it examined the feasibility of the test in terms of time requirement and mechanics of implementation. It also ascertained the clarity of instructions and test items, as well as typographical errors and other relevant concerns. The result of the first pilot test was a requisite for the test item analysis.
4. **First Item Analysis** – The first item analysis examined the quality (validity) of individual items. Among the approaches of item analysis, the study established the indices of difficulty and discrimination as well as distractor efficiency. These approaches guided the researchers to determine the items to be retained, revised or discarded. Items with acceptable difficulty index, discrimination index and distractor efficiency were selected for inclusion in the second draft of the test.
5. **Revisions after the First Pilot test** – Guided by the results of the item analysis, the same test constructors refined the test items. Only potentially good items that cover particular competencies were considered. Hence, the test items

included in the final form of the CRT were based on the following criteria: (a) each competency should have a representation of at least two items in order to confirm mastery of the competency; (b) items with moderate difficulty index and good discriminatory power are prioritized. Items that have good difficulty index but poor discriminatory power were revised (either the stem or options as required). Of the 300 original items in the first pilot test, only 200 items were retained.

6. Second Pilot Test – The second pilot test was conducted to a bigger sample size in the three (3) congressional districts of the select province. This phase further examined the quality of test items in terms of their validity and reliability.

7. Second Item Analysis and Reliability Test – The second item analysis followed the same procedures and criteria in the first item analysis. This phase further refined the final 200 test items of the CRT. Moreover, internal consistency of the CRT was established using Kuder- Richardson Formula 20.

8. Revisions after the Second Pilot test – The revision was focused on improving items having poor distractors, modifying the stems, and options of items which were not gender-fair, and enhancing the pictures and graphics for better visibility.

9. Finalization of the CRT – The CRT was subjected to proofreading, lay-outing, and designing by experts before its reproduction. A manual was also developed to document the processes it underwent and to serve as a guide for its administration, scoring, and interpretation. The final print-out was submitted for copyright registration to the National Library of the Philippines.

3.5 Ethical Considerations

The researchers sought the approval of the concerned school authorities in the conduct of the study. They explained personally the intent of the research as well as the expected outcome. Free, prior, and informed consent was also obtained to each respondent from the different schools in the select province. Moreover, the respondents understood well the overview of the research. They knew that their participation was voluntary and that any information obtained from them would be treated with utmost confidentiality.

3.6 Data Analysis

The descriptive statistics such as mean, standard deviation, as well as frequency count and percentage were used for the item analysis. In particular, item difficulty was calculated using the following formula of Crocker & Algina (1986):

Difficulty index = (Number who answered an item correctly divided by the overall number of tested) multiplied by 100

The item difficulty index was determined using Gregory's (2000) indices of difficulty as follows: 14% and below, Very Difficult; 15% - 39%, Difficult; 40%-70%, Desirable; 71% - 85%, Easy; and 86 - 100% Very Easy.

On one hand, the point-biserial correlation was used in obtaining the discrimination coefficient. This statistical tool is viewed as the most efficient means of examining item discrimination because it includes all examinees in the analysis (Essen & Akpan, 2018; McCowan & McCowan, 1999). The following criteria were used for interpreting the correlation: (a) less than +0.09, poor; between +0.09 and +0.30, reasonably good; greater than +0.30, very good items (Educational Assessment Corporation, 2009).

In terms of the distractor efficiency, the researchers adopted the criteria from the study of Mahjabeen et al., (2017). If <5% of the respondents selected the incorrect answers, the distractors are regarded as non-functional distractor (NFD) whereas if >5% of the respondents selected the incorrect answers, the distractors are known as functional distractors (FD). The classification and interpretation of distractor efficiency are as follows:

Type of Non-functional Distractors	Interpretation
0 Non-Functional Distractors	100% Distractor Efficiency
1 Non-Functional Distractors	66.6% Distractor Efficiency
2 Non-Functional Distractors	33.3% Distractor Efficiency
3 Non-Functional Distractors	0% Distractor Efficiency

Lastly, Chi-Square was employed in ascertaining the association between CRT quality indices whereas Kuder-Richardson Formula 20 was used to determine the reliability (inter-item consistency) of the CRT. A test with computed value greater than .70 is considered reliable (Crocker & Algina, 2008; Smith, 2018).

4. Results and Discussion

4.1 Item Analysis of the First Pilot Test

Table 1. Characteristics of the CRT items based on the evaluation criteria

<i>Parameters</i>	<i>Result</i>
Total number of examinees	231
Total number of CRT items	300
Total Score	300
Score Obtained	
Mean \pm SD	99.10 \pm 19.82
Range	79.28 – 118.93
Difficulty Indices	
Mean \pm SD	37.13 \pm 16.00
Range	21.11 – 53.13
Discrimination Indices	
Mean \pm SD	0.20 \pm 0.15
Range	0.04 – 0.35
Distractor Efficiency	
Mean \pm SD	98.19 \pm 9.45
Range	0 – 100

Table 1 shows that 231 examinees took the 300-item CRT for its first pilot test. The test items were generally *difficult* as the computed mean of the index of difficulty was 37.13 (range=21.11-53.13). Out of the 300 items and with a statistical range of 79.28 to 118.93, the estimated scores ranged the lowest at 79 to the highest at 119. This finding suggests that some of the test items were not ideal thus, they needed refinement and revisions.

Conversely, most of the items were within the acceptable parameters of *reasonably good* as revealed by the computed mean discrimination index of 0.20 (range= 0.04 - 0.35). This means that most of the test items were able to discriminate academically the high performing and low performing examinees. This finding is consistent with the computed distractor efficiency of 98.19 (range= 0 - 100%) revealing that nearly all of the items had functional distractors (see also Table 2).

Table 2. Categories of distractors by distractor efficiency

<i>CRT Items with Non-functional Distractors</i>	<i>Frequency</i>	<i>Percent</i>	<i>Distractor Efficiency (%)</i>
0 Non-Functional Distractors	288	96	100
1 Non-Functional Distractors	9	3	66.6
2 Non-Functional Distractors	3	1	33.3
3 Non-Functional Distractors	0	0	0
Total	300	100	
Number of Distractors	900		

Table 2 reveals that the 300-item CRT has 900 distractors. Remarkably, the test items possessed 100% distractor efficiency since 288 items (96%) had zero non-functional distractors. Nine (9) items (3%) had 1 non-functional

distractors or 66.6% distractor efficiency whereas only 3 items (1%) had 2 non-functional distractors or 33.3% distractor efficiency. Interestingly, none of the items reflected 0% distractor efficiency.

Moreover, the 300-item CRT possessed a great number of functional distractors and it was consistent with the finding that the distractors were good. However, a careful inspection of the results of item analysis showed that guessing of some items was probable because several examinees both from the upper and lower groups chose one of the distractors. Such concern may be attributed to the fact that most of the items were difficult (see Tables 1 & 3). Hence, this finding became an essential input for the revision of some test items.

Table 3. Cross-tabulation of the index of difficulty and index of discrimination

<i>Difficulty Index</i>		<i>Discrimination Index</i>			<i>Total</i>
		<i>Poor</i>	<i>Reasonably Good</i>	<i>Very Good</i>	
Difficult	n	68	96	19	183
	%	22.67	32.00	6.33	61.00
Desirable	n	7	31	59	97
	%	2.33	10.33	19.67	32.33
Easy	n	1	4	7	12
	%	0.33	1.33	2.33	4.00
Very Easy	n	4	4	0	8
	%	1.33	1.33	0.00	2.67
N		80	135	85	300
%		26.67	45.00	28.33	100

Table 3 presents that of the 300 items, 183 (61%) were *difficult*. Only 97 items (32.33%) were assessed to be *desirable*, 12 items (4.0%) were *easy*, and 8 items (2.67%) were *very easy*. This finding means that the examinees found the CRT to be generally challenging, which is consistent with the computed index of difficulty (see Table 1). This finding could be attributed to the fact that as a criterion-referenced test, 70% of the items were based on the performance standards of the CRS and only 30% of the items were based on the content standards.

Moreover, using point biserial coefficients, 135 items (45%) were found to be *reasonably good* and 85 items (28.33%) were *very good*. A total of 80 items were considered *poor*, which were then ultimately rejected. To further improve the test items, the researchers and the expert validators revisited the difficulty and discrimination indices of each item. This procedure enabled them to drop additional 20 items making the CRT a 200-item test ready for the second phase of the pilot test. This move reduced the examination length without compromising the set of competencies to be measured. According to Foley (2016), reducing the test length may also lessen the test anxiety and burden of the examinees and can generate acceptable assessment design choices.

4.2 Item Analysis of the Second Pilot Test

Table 4. Characteristics of the CRT based on the evaluation criteria

Parameters	Result
Total number of examinees	942
Total Number of CRT Items	200
Total Score	200
Score Obtained	
Mean \pm SD	129.32 \pm 13.51
Range	115.81 – 142.83
Difficulty Indices	
Mean \pm SD	65.54 \pm 22.43
Range	43.11 – 87.97
Discrimination Indices	
Mean \pm SD	0.22 \pm 0.13
Range	0.09 – 0.35
Distractor Efficiency	
Mean \pm SD	68.91 \pm 33.64
Range	0 – 100

Table 4 shows that 942 examinees participated in the second pilot test. The test consisted of 200 items and the estimated scores ranged from 115.81 to 142.83 implying that the lowest score expected is 116 and the highest score expected is 143. Results revealed that the CRT test items were *desirable* as shown by the mean of 65.54 and the estimated computed indices of difficulty ranged from 43.11 to 87.97. This finding is consistent with the claim of Sahoo and Singh (2017) positing that an ideal item has a difficulty index between 30% and 70%. In a criterion-referenced test like the CRT, moderate difficulty is ideal because its focus is on the mastery of the competencies (Berk, 1986). Moreover, items of criterion-referenced tests are usually distributed across various levels of difficulty since learning outcomes in the educational context span from the simplest to the most complex in consideration of all the cognitive tasks (Bond, 1996).

Using point-biserial coefficients measures, the CRT items were *reasonably good* considering that the estimated discrimination indices ranged from 0.09 to 0.35 (mean=0.22). This finding indicates that the CRT is a good tool in differentiating examinees vis-à-vis their level of college readiness as being measured by the test items. Those students who got high scores in the test responded to most items correctly, while those who obtained low scores answered most of the items incorrectly.

Table 5. Categories of distractors by distractor efficiency

CRT Items with Non-Functional Distractors	Frequency	Percent	Distractor Efficiency
0 Non-Functional Distractors	88	44.0	100
1 Non-Functional Distractors	60	30.0	66.6
2 Non-Functional Distractors	31	15.5	33.3
3 Non-Functional Distractors	21	10.5	0
Total	200	100	
Number of distractors	600		

It can be gleaned from Table 5 that the 200-item CRT has 600 distractors. Of this number, 88 items (44%) possessed 100% distractor efficiency. Sixty (60) items (30%) had 1 non-functional distractors or 66.6% distractor efficiency while 31 items (15.5%) had 2 non-functional distractors or 33.3% distractor efficiency. Only 21 items (10.5%) had 3 non-functional distractors or 0 distractor efficiency because they did not attract any of the examinees. The data reveals that generally, the CRT has large number of functioning distractors, thus possessing the acceptable distractor efficiency (DE). As claimed by Haladyna, Downing, & Rodriguez (2002), “the greater the number of plausible distractors, the more accurate, valid, and reliable the test typically becomes.”

Table 6. Association between the CRT quality indices

<i>Indices</i>	<i>df</i>	<i>Chi-square Value</i>	<i>p-value</i>
Difficulty Index and Discrimination Index	6	30.375	0.00
Difficulty Index and Distractor Efficiency	9	152.486	0.00
Discrimination Index and Distractor Efficiency	6	6.838	0.34

Using chi-square to test the relationship between the indices of difficulty and discrimination, the two measures were significantly dependent on each other ($X^2=30.375$; $p=0.00$). The positive correlation establishes that the more difficult the CRT items are, the more they discriminate high and low performing examinees. In short, difficult items of the CRT are able to discriminate high and low performing examinees. High performing examinees are likely to answer the difficult items correctly, more than the low performing ones. As claimed by previous studies, moderately challenging items demonstrate good discriminating potential (Boopathiraj & Chellamani, 2013; Sim & Rasiah, 2006). Corroboratively, a strong association between discrimination and difficulty indices shows that test items possess quality and credibility (Suruchi and Rana, 2014).

Significantly, results revealed a strong association between distractor efficiency and difficulty index of the CRT items ($x^2=152.486$; $p=0.00$). This finding means that the CRT items that have higher distractor efficiency have higher difficulty index. In other words, the more efficient the distractors are, the more difficult it would be for the less-prepared examinee to discern the correct answer. As Hingorjo & Jaleel (2012) claimed, test items with options that efficiently attract examinees to select them tend to be more challenging.

4.3 Internal Consistency of the CRT

To establish the CRT’s reliability, the inter-item consistency of the test was tested. Using Kuder-Richardson Formula 20, the computed value was $r=0.796$ implying that the reliability is within the acceptable range of 0.70 - 0.90 (Crocker & Algina, 2008). Examinees who successfully answered a difficult test item were also able to answer correctly the easy ones. Thus, the designed and validated CRT can provide consistent results when used in measuring college readiness of Filipino K to 12 graduates. As asserted by Neuman (2003), one primary indicator of a good test is that when it is applied to the same object of measurement several times, it can yield similar results.

4.4 Contextualized and Gender-Fair CRT Items

Contextualization is one important feature of the CRT. According to Berns and Erickson (2001), contextualization ensures the teaching of skills with “direct reference to real world events and practices”. In fact, making examination questions fit in real-world context has become a prevalent practice in the academe (Ahmed & Pollit, 2007). As affirmed by Perin (2011), it is important that activities, issues, and authentic materials related to the learners must be used in meeting their needs. With this, the CRT has items that are considered meaningful and relevant to the context of the test takers that may enable them to build their confidence in taking the test. Samples of contextualized test items are shown in the following figures. Figure 2 shows that the test items were contextualized since the focus of the reading selection was one of the pressing concerns in the locale of the study. This issue had engendered debates and varied views from the people. The selection was published in a local newspaper and was circulated through the internet, too.

For items 20 – 22, read carefully the article and answer the questions that follow.

Black sand or magnetite mining is being done now in almost every part of the country. In Cagayan, this is an ongoing problem, with both small scale and large scale miners coming in to gather black sand from both on- and offshore. This is being allowed by the local governments and other executive offices despite the Department of the Interior and Local Government (DILG) Memo Circular No. 44, s. 2014 clarifications that foreshore mining is not allowed and that there is still a need for clear guidelines for offshore mining.

Black sand mining has been opposed by Church and peoples' organizations since the beginning. Initial research shows that magnetite mining causes agricultural and environmental problems. Black sand mining in coastal areas result in coastal erosion, inundation of communities, and degradation of marine ecosystem.

In 2009, ~~Alvansa Tigil~~ ~~Mina~~ (ATM) sent a team to look into the reported beach erosion in Cagayan province, reportedly due to the illegal black sand mining operations in the area. In 2010, an anti-mining leader was killed.

Source: <https://ejatlas.org/conflict/cagayan-black-sand-mining-in-shore-area>
(Retrieved: February 12, 2019)

20. The first paragraph claims that black sand or magnetite mining is rampantly done in every part of the country. Is the statement true?
A. Yes
B. No
C. Not stated
D. Undecided

21. What is the writer's purpose in sharing this article?
A. Encourage all religious groups to stop illegal acts
B. Establish better relationship with miners and local officials
C. Develop concern and awareness of the effects of black sand mining
D. Apply the guidelines stipulated in DILG Memo Circular No. 44, s. 2014

Figure 2. Sample items in English

Figure 3 also shows the feature of contextualization particularly in Social Studies learning. Item no. 146 highlighted the recent archeological discovery in the select province while item no. 147 focused on a national contemporary issue that was passionately discussed by the people and televised by major meadia outlets in the Philippines. These items made the test to be localized, one important feature of contextualization.

146. Where was the fossil of the oldest man in the Philippines found?
A. ~~Bolinao, Pangasinan~~
B. ~~Peñablanca, Cagayan~~
C. Quezon, Palawan
D. Rizal, Kalinga

147. Which of the following bells served as war trophies during the Philippine-American War and was recently returned to the Philippines in 2018?
A. Cebu City, bells
B. ~~Bauang, La Union~~ bells
C. ~~Camalaniugan, Cagayan~~ bells
D. ~~Balangiga, Eastern Samar~~ bells

Figure 3. Sample items in Social Studies

Besides contextualization, the CRT items were also designed to be gender-fair. The use of gender fair language has the potential to make significant contributions to the reduction of gender stereotyping and discrimination (Sczesny, Formanowicz, & Moser, 2016). This process was done to ensure that CRT items do not carry any gender bias. Figure 4 shows the usage of words in the test items that are gender-inclusive (use of his or her as a pronoun) while Figure 5 presents the usage of words that are gender-neutral (use of one or a person as a subject) as a way of treating women and men symmetrically.

11. When listening to a speech, you can identify the speaker's main idea through his/her?
- A. Detailed anecdotes and reports
 - B. Words and phrases that are repeated
 - C. Words that indicate related ideas to the main topic
 - D. Emphasis on the most important points in his/her delivery
33. "The character does not change overtime. His or her personality does not transform or evolve." Which type of character is depicted in the preceding statement?
- A. Dynamic
 - B. Foil
 - C. Static
 - D. Stock

Figure 4. Sample items in English (item no. 11) and Literature (item no. 33)

120. In travelling from a forest ecosystem to a grassland, one observes that the trees gradually give way to prairie grasses. What critical factor governs this shift in the ecosystem?
- A. Availability of water
 - B. Change in elevation
 - C. Annual mean temperature
 - D. Length of the growing season
121. Which of the following factors determines a person's athletic ability?
- A. Genetic coding
 - B. Parental influences
 - C. Environmental factors
 - D. Genetic and environmental factors

Figure 5. Sample items in Science

4.5 Interpreting the CRT

A college ready student is one who will obtain a raw score of 100 or more out of the 200 items of the CRT. A 100 raw score has an equivalent percentage of 75%. In the Philippine basic and higher education institutions, generally, 75% is the passing rate in major examinations. Considering that the CRT is a criterion-reference instrument, the focus is to measure the ability of the would-be college students to master the competencies aligned with the Philippine CRS as indicator of college readiness.

5. Limitations of the Study

The validity and reliability of the CRT as a measure of college readiness is limited to the following contexts and factors. First, the designed and validated CRT, as a new construct, is just an initial attempt to develop a standardized tool measuring college readiness of the K to 12 graduates based on the Philippine context. As it is a work in progress, its reliability and validity can be further tested to other higher education institutions in the Philippines for greater generalizability. Second, the CRT measures only the college readiness of Filipino K to 12 graduates to hurdle the general education courses in college as defined in the CRS. It does not measure their readiness to be admitted and to succeed in a specific program in the tertiary level. Third, the validity and reliability of the CRT have limited applicability considering that the sampling and pilot testing were just limited to one geographical province in the Philippines. Thus, its results hold true only to its locale. Fourth, the CRT is only limited to 200 items measuring the examinees' ability to demonstrate knowledge and skills defined in the CRS. As it is content-focused, the items cover content and performance competencies.

6. Conclusion and Recommendations

The 200-item CRT is a valid and reliable instrument to measure the college readiness of Filipino K to 12 graduates with its features of being criterion-referenced, contextualized, and gender-fair. It has the ability to identify K to 12 graduates who could be admitted in college and who could pass without remediation the general education courses.

Given the findings of the study, the following actions are recommended:

1. The CRT must be validated at the national level for greater generalizability and it must be further subjected to other validity and reliability measures.
2. Further validity of the CRT must be tested to examine whether it can predict the college readiness of the K to 12 graduates to succeed in a specific program in the tertiary level and not on general education courses only.
3. The concurrent validity of the CRT must be tested with the College Admission Test result and other relevant tests that measure transition of K to 12 graduates from basic education to tertiary level.
4. The CRT may be utilized by public and private basic education to determine the college readiness of the Philippine K to 12 graduates.
5. The CRT may be utilized by Philippine public and private Higher Education Institution (HEIs) as a college admission test or as a complement or a substitute for their current admission test.

Acknowledgements

The data of the study emanated from the research project funded by the Philippine Commission on Higher Education (CHED) under the Discovery-Applied Research and Extension for Trans/Inter-disciplinary Opportunities (DARE TO) Grant (Commission En Banc Resolution No. 187-2019). The researchers are grateful to all the student-respondents, expert validators, and test constructors from the Department of Education (DepEd) and the Cagayan State University (CSU). Special thanks are given to Prof. Marcelo Raquepo for his technical assistance as well as to Commissioner Lilian de las Llagas of CHED and President Urdujah Tejada of CSU for their effective mentoring.

References

- Adelman, C. (2006). *The toolbox revisited: paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education. <https://eric.ed.gov/?id=ED490195>
- Ahmed, A. & Pollit, A. (2007). Improving the quality of contextualized questions: an experimental investigation of focus. *Assessment in Education: Principles, Policy & Practice*, 14(2), 201-302. <https://doi.org/10.1080/09695940701478909>
- Avery, C. & Kane, T.J. (2004). *Student perceptions of college opportunities: The Boston COACH program*. In C. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 355-394). Chicago, IL: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226355375.003.0009>
- Barnes, W., Slate, J. & Rojas-LeBouef, A. (2010). College readiness and academic preparedness: The concepts? *Current Issues in Education*, 13(4), 1-28. Retrieved from <http://cie.asu.edu/>
- Barnes, W. & Slate, J. (2013). College-readiness is not one-size-fits-all. *Current Issues in Education*, 16(1), 1-13.
- Berk, R. (1986). Guide to setting performance standards on criterion-referenced tests. *Review of Educational Research Spring*, 56(1), 137-172. <https://doi.org/10.3102/00346543056001137>
- Berns, R. G. & Erickson, P. M. (2001). *Contextual teaching and learning: Preparing students for the new economy* (The Highlight Zone: Research @ Work No. 5). Louisville, KY: University of Louisville, National Research Center for Career and Technical Education.
- Bond, L. (1996). Norm-and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 5(2). 1-3.
- Boopathiraj, C. & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189-193.
- Bridgeman, B. (1991). Essays and multiple-choice tests as predictors of college freshman GPA. *Research in Higher Education*, 32, 319-331. <https://doi.org/10.1007/BF00992895>
- Burton, N. W. & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1990*. College Board Research Report 2001-2. Retrieved from http://professionals.collegeboard.com/profdownload/pdf/rdreport200_3919.pdf

- Camara, W. J. & Echternacht, G. (2000). *The SAT I and high school grades: Utility in predicting success in college* (Research Notes RN-10). The College Board, Office of Research and Development. <http://research.collegeboard.org/sites/default/files/publications/2012/7>
- CEB Resolution No. 298-2011. College Readiness Standards. Retrieved from http://www.digigogy.com/uploads/4/5/6/7/4567766/cmo-20s-2013_appendix_c.pdf
- CMO No. 10, Series of 2017. Retrieved from <https://ched.gov.ph/cmo-10-s-2017-2/>
- Conley, D. T. (2007). *Toward a more comprehensive conception of college readiness*. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T. (2010). *College and career ready: Helping all students succeed beyond high school*. San Francisco, CA: Jossey-Bass. <https://doi.org/10.1002/9781118269411>
- Conley, D. T., Aspengren, K., Stout, O. & Veach, D. (2006). *College Board Advanced Placement best practices course study report*. Eugene, OR: Educational Policy Improvement Center.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Crocker, L. & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.
- Dougherty, C., Mellor, L. & Jian, S. (2006). *The Relationship between advanced placement and college graduation*. Washington, DC: National Center for Educational Accountability.
- Educational Assessment Corporation (2009). Retrieved July 3, 2019 from <https://www.edassess.net/eacs/pointbiserial.aspx>.
- Essen, C. & Akpan, G. (2018). Analysis of difficulty and point-biserial correlation indices of 2014 Akwa Ibom State Mock Multiple Choices Mathematics Test. *International Journal of Education and Evaluation*, 4(5), 1-11.
- Foley, B. (2016). Getting lucky: How guessing threatens the validity of performance classifications. *Practical Assessment, Research & Evaluation*, 21(3). 1-23. <http://pareonline.net/getvn.asp?v=21&n=3>
- Fraenkel, J. R. & Wallen, N. E. (2003). *How to design and evaluate research in education* (5th ed.). New York, NY: McGraw-Hill Higher Education.
- Greene, J. P. & Forster, G. (2003). *Public high school graduation and college readiness rates in the United States*. New York, NY: Center for Civic Innovation, Manhattan Institute.
- Gregory, R. (2000). *Psychological testing: History, principles, and applications* (3rd ed.). Needham Heights, MA: Allyn & Bacon, Inc.
- Haladyna, T.M., Downing, S.M. & Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334. https://doi.org/10.1207/S15324818AME1503_5
- Hingorjo, M.R. & Jaleel, F. (2012). Analysis of One-Best MCQs: the difficulty index, discrimination index and distractor efficiency. *Journal of the Pakistan Medical Association*, 62(2). 142-147.
- ICEF Monitor. (2013). *Philippines creates opportunities in overhaul of K-12 education system*. Retrieved from <https://monitor.icef.com/2013/08/philippines-creates-opportunities-in-overhaul-of-k-12-education-system/>
- Jawad, L. (2017). *Examining college readiness in an early college program that focuses on health careers: Perceptions of graduating students*. Dissertation presented at the University of Michigan-Dearborn.
- Kane, M. (2001). *So much remains the same: Conception and status of validation in setting standards*. In G. Cizek (Ed.), *Setting performance standards* (pp. 53-88). New York, NY: Routledge.
- Kless, L., Soland, J. & Santiago, M. (2013). *Analyzing evidence of college readiness: A tri-level empirical and conceptual framework*. Baltimore, MD: John W. Gardner Center for Youth and Their Communities. Retrieved from <http://gardnercenter.stanford.edu/resources/publications/collegereadinesslitreview.pdf>
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D. & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average* (College Board Research Report No. 2008-5). New York: The College Board. Retrieved June 6, 2019 from http://professionals.collegeboard.com/profdownload/pdf/08-1718_RDRR_081017_Web.pdf
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D. & Barbuti, S. M. (2008). *Validity of the SAT for predicting*

- first year college grade point average*. New York, NY: College Board.
- Landmark College (2009). *Landmark College's guide to assessing college readiness for parents of college-bound children with learning disabilities or AD/HD*. Putney, Vermont: Landmark College.
- Leonard, J. (2013). Maximizing college readiness for all through parental support. *School Community Journal*, 23(1), 183-202. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1004338.pdf>
- Magno, C. & Piosang, T. (2016). Assessment schemes in the senior high school in the Philippine basic education. *Educational Measurement and Evaluation Review*, 7, 66-87.
- Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S. & Rizvi, M. (2017). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of Pakistan Institute of Medical Science*, 13(4), 310-315.
- Mateo, J. (2019, February 7). Grade 12 students to undergo exit assessment. *The Philippine Star*. Retrieved from <https://www.philstar.com/other-sections/education-and-home/2019/02/07/1891485/grade-12-students-undergo-exit-assessment#IKIZQIZfzCxMAce.99>
- McCowan, R. & McCowan, S. (1999). *Item analysis for criterion-referenced test*. New York: The Center for Development of Human Services.
- McMillan, J. H. & Schumacher, S. (2001). *Research in education: A conceptual introduction* (5th ed.). New York: Longman.
- Merisotis, J. & Phipps, R. (2000). Remedial education in colleges and universities: What's really going on? *The Review of Higher Education*, 24(1), 67-85. <https://doi.org/10.1353/rhe.2000.0023>
- Messick, S. (1989). *Validity*. In R. L. Linn (Ed.). *Educational measurement* (3rd Ed.). New York: American Council on Education and Macmillan.
- Moss, P. (1994). Can there be validity without reliability? *Educational Researcher*, 23, 5-12. <https://doi.org/10.3102/0013189X023002005>
- Neuman, W.L. (2003) *Social Research Methods: Qualitative and Quantitative Approaches*. New York: Allyn and Bacon.
- Okabe, M (2013). *Where does Philippine education go? The "K to 12" program and reform of Philippine basic education*. IDE Discussion Paper No. 425. Japan: Institute of Developing Economies.
- Olson, L. (2006, April 26). Views differ on defining college prep: How to gauge readiness sparks vexing questions. *Education Week*, 25(33), 28-29.
- Perin, D. (2011). *Facilitating student learning through contextualization*. CCRC Working Paper No. 29. Community College Research Center. <https://doi.org/10.1177/0091552111416227>
- Porter, A. & Polikoff, M. (2011). Measuring academic readiness for college. *Educational Policy*, 26(3), 394-417. <https://doi.org/10.1177/0895904811400410>
- Republic Act No. 10533 (2013). Enhanced Basic Education Act of 2013. Retrieved May 7, 2019, from <https://www.officialgazette.gov.ph/2013/05/15/republic-act-no-10533/>
- Reuschel, J. (2009). *A Comparative study of high school academic paths, grade point averages, and ACT composite scores as predictors of success at Walters State Community College*. Electronic Theses and Dissertations. Paper 1847. <https://dc.etsu.edu/etd/1847>
- Roderick, M. (2006). *Closing the aspirations-attainment gap: Implications for high school reform*. New York, NY: MDRC High School Reform.
- Sahoo, D.P. & Singh, R. (2017). Item and distractor analysis of multiple choices questions (MCQs) from a preliminary examination of undergraduate medical students. *International Journal of Research in Medical Sciences*, 5(12), 5351-5355. <https://doi.org/10.18203/2320-6012.ijrms20175453>
- Sczesny, S., Formanowicz, M. & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, 7(25), 1-10. <https://doi.org/10.3389/fpsyg.2016.00025>
- Sim, S.M. & Rasiah, R.I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper, *Annals of the Academy of Medicine of Singapore*, 35(2), 67-71.

- Smith, S. (2018). What is KR20? *Platinum Educational Group*. Retrieved from <https://platinumed.zendesk.com/hc/en-us/articles/214600006-What-is-KR20->
- Suruchi & Rana, S. (2014). Test item analysis and relationship between difficulty level and discrimination index of test items in an achievement test in Biology. *Indian Journal of Research*, 3(6), 56-58. <https://doi.org/10.15373/22501991/June2014/18>
- Taghoy, E. (2019, June 10). CHED clarifies issues on college admission policy. *Philippine Information Agency*. Retrieved from <https://pia.gov.ph/news/articles/1023003>