

Feasibility Evaluation of Aptitude Testing in Civil Service Examinations (English Version) *

Po-Hsi Chen**

Abstract

This article investigates the civil service examination from multiple perspectives. The author analyzes related factors and considerations when applying the aptitude testing in civil service examination. Aptitude testing has been used in military and industry for personnel and employee selection since the First World War. Progress in statistical methods, the developing trend in intelligence theories and the tendency of the college entrance system lay the foundation for using the multiple latent traits in civil service examinations. Aptitude testing is the best tool for evaluating multiple talent. The Ministry of Examination (MOEX) has carried out several projects applying different psychological tests in civil service examinations. To realize the ideal of using aptitude testing in civil service examinations, the author suggests five factors for practical consideration: human, administrative, time, space, and tools. Evaluation of these factors can be for reference in policy making and practical planning for using aptitude testing in civil service examinations.

Keywords: aptitude testing, civil service examination, feasibility evaluation

* This article was written by Professor Chen and translated into English by the Ministry of Examination

** Professor, Department of Educational Psychology and Counseling, National Taiwan Normal University

公務人員考試加入性向測驗的可行性評估 (英譯版) *

陳柏熹*

摘要

本文主要是以多元觀點來探討公務人員考試的選才方式，並分析在公務人員考試中若加入性向測驗作為選才參考依據之一，需要考量哪些因素。測量多元潛能的性向測驗，早在第一次世界大戰前就普遍被用在軍事及工業中的人才甄選與職涯諮商，不論是統計方法的演進，還是智力理論的取向轉變，乃至於學校的入學選才，都為多元潛能在公務人員考試的應用奠定了良好的基礎，而性向測驗的主要目的正是用來評估應試者的多元潛能。考選部近二十年來陸續對於將不同考試方式納入公務人員考試的選才中作了法規制度的改變、評估研究案，以及試題範例的研發，但真正要將不同測驗方式落實公務人員考試中，還需要從人、事、時、地、物等五個層面作仔細的考量與及規劃。本文從這些層面進行了現況分析，以作為未來公務人員考試要將性向測驗納入公務人員選才時的政策擬訂及實務規劃參考。

關鍵詞：性向測驗、公務人員考試、可行性評估

* 本文由陳柏熹教授撰文，考選部翻譯為英文版

**國立臺灣師範大學教育心理與輔導學系教授

I. Introduction

It has been more than a hundred years since the first standardized psychometric testing Binet-Simon Intelligence Scale (Binet & Simon, 1905) was developed. While the scale provides a standardized procedure to assess general mental abilities, psychologists quickly found the limitations of a test designed to measure verbal and numerical skills. They soon began developing special aptitude tests to compensate for shortcomings of the general intelligence test, measured subjects' multiple latent traits, and catered to the hiring needs of large-scale industrial and military recruitment for different types of abilities.

A. Development of the Multiple Aptitude Batteries

The aptitude test is an example of a maximum performance test. It has answers that are right or wrong and is designed to assess an individual's potential (Anastasi & Urbina, 1997). A multiple aptitude test, also known as multiple aptitude battery, is designed to measure special aptitudes. The evolution of multiple aptitude batteries is associated with the following trends:

(A) Theories of Intelligence

The definition and measure of intelligence has been much discussed in psychology. In a 1921 issue of the *Journal of Educational Psychology*, scholars discussed intelligence and its measurement. Although the Stanford-Binet Intelligence Scale proposed the idea of intelligence quotient (Terman, 1916) and simplified it with ratio between biological and mental age, many psychologists disagreed with the idea of representing intelligence by a number and shared their different views. For example, Spearman's two factor theory proposes that mental processes require both general and specific factors and that an individual's performance on tests or tasks is determined by these two types of factors (Spearman, 1927). The wider the gap between test types or task activities, the smaller the influence of general factors and the greater the influence of specific factors.

As Spearman did not elaborate further on the types of specific factors, the need for a multi-factor theory arose. Thurstone (1938) is the representative of the multiple-factor theory. Take the intelligence theory proposed by Thurstone and Thurstone (1941) for example. They held that human's primary mental abilities include verbal comprehension, word fluency, numeric ability, spatial relations, associative memory, perceptual speed, and inductive reasoning. Other scholars believed that intelligence should be viewed as consisting of three dimensions: operations, contents, and products. This theory is known as the Structure-of-Intellect model, proposed by Guilford (Guilford, 1988).

One of the best-known contemporary views on multiple latent traits is Gardner's theory of multiple intelligences (Gardner, 1983). This theory divides human intelligence into eight types: linguistic, visual-spatial, musical, bodily-kinesthetic, logical-mathematical, naturalistic, interpersonal, and intrapersonal. Today, we can find traces of Gardner's theory in Taiwan's multiple admission system. Gardner serves as senior director of Project Zero with government and business support and uses assessment centers to observe and record the development of multiple intelligences and offers training courses.

These intelligence theories support one idea: A single score or a single ability is not sufficient to explain an individual's performance on different task activities or tests. Human abilities should, rather, be viewed through multiple latent traits and multiple assessments.

(B) Development and Applications of Factor Analysis

The concept of factor analysis was first introduced by Spearman (1927) to identify related factors. Later, Thurstone (1931) proposed multiple factor analysis. Factor analysis makes important contributions to understanding the abilities required for test construction and task activities and led to the development of various types of multiple latent traits in the abovementioned intelligence theories.

Multiple aptitude battery tests assess an individual's different potentials. They allow test results to be represented not merely by a single IQ score but with a profile diagram that shows an individual's ability in areas such as language, numbers, spatial perception, and perceptual speed. These factor-analysis-based tests were used mainly for military purposes, especially during World War II, when multiple aptitude batteries were used to recruit military personnel with various abilities. These tests could compensate for the limitations of single-ability tests like Army Alpha and Army Beta. This trend led to the development of Armed Services Vocational Aptitude Battery (ASVAB), which to this day remains the primary enlistment test for the US armed forces. Currently, the ASVAB can generate more than 50 different combinations of measurements to meet the requirements of different military positions (ASVAB, 2023).

(C) Trends in School Admissions

The earliest use of the Binet-Simon Intelligence Scale was by schools in France to screen children with mental retardation, not to assess learning outcomes. The first standardized testing to evaluate learning outcomes was the Stanford Achievement Test, published in 1923. The test used a single norm-referenced group to present learning outcomes of students from different schools. Then, with the establishment of College Entrance Examination Board (CEEB) and Educational Testing Service (ETS),

achievement tests evolved to measure academic aptitude of students. The Scholastic Aptitude Test (SAT) was introduced under this trend. The development of the SAT was similar to that of intelligence theories. The SAT came with a single core in its 1926 debut. In 1931, it evolved to include verbal and mathematical components. In 1993, it developed into SAT I and SAT II, with the former assessing reasoning abilities and the latter assessing specific subjects. The SAT generally referred to is SAT I, which is an aptitude test used to predict future academic success and not an achievement test.

The major difference between aptitude and achievement tests is that aptitude tests are administered before learning to evaluate an individual's ability to learn or potential to perform tasks. Aptitude test scores serve as a reference for the individual to make informed decisions between academic or vocational careers. On the other hand, achievement tests are administered after learning to evaluate an individual's learning outcomes. Aptitude tests can be divided into academic and vocational types. Academic aptitude tests focus on language, math, and logical reasoning—skill areas most associated with learning. Vocational aptitude tests measure a wider variety of skills, including the abovementioned traits but hands-on areas closely associated with workplace activities such as perceptual speed, mechanical aptitude, and visual-spatial awareness.

In Taiwan, university admission started using a single-admission-method system, the Joint College Entrance Examination in 1954. It was not until 2001 that multiple admission pathways were used. Schools can form their own admission policy. Today, the three admission methods in use are: the Star Admission Program (5~10%), referral application admission (40~50%), and test score based placement (40~50%). The test score based placement is based solely on test scores. The other two methods consider comprehensive learning outcomes, including learning portfolios (school learning outcomes and multifaceted learning performance) and may incorporate other admission processes like oral, practical, and written examinations (Lee et al., 2012; Yeh & Ting, 2021).

The core abilities or essential competences schools want to foster in their students also follow a multiple latent trait approach. Chen et al. (2015) compiled the core abilities and essential competences published by roughly a dozen public and private universities and colleges in Taiwan and sought reference from the Assessment & Teaching of 21st Century Skills[ATC21S] (2010). They came up with nine essential competencies for college students: communication and teamwork, aesthetic literacy, scientific reasoning, information literacy, lifelong learning, innovative leadership, problem-solving, civil society awareness, and career development. The researchers developed a basic competence test for college students to measure the nine competences with different question types. The development also suggests that the idea of a single

test score ill fits the modern concept of developmentally appropriate practice. Schools are interested in knowing if applicants possess the various qualities required for learning and fostering the core abilities needed for their career.

(D) Public and Private Sector Recruitment Trends

Many examples of the public sector incorporating aptitude tests into recruitment process can be found around the world. In Asia, both the Public Service Division and Monetary Authority of Singapore use psychometric testing (including aptitude tests) as part of the selection process. The Japanese civil service examination uses the Uchida-Kraepelin psychodiagnostic test, a kind of aptitude test designed to evaluate math skills and concentration. The US's Office of Personnel Management[OPM] (2019) uses intelligence and aptitude tests to help screen federal officials. State governments also use personality, aptitude, and intelligence tests as evaluation tools for selection, elimination, and promotion. Similar practices can be found in European countries, including the UK and Germany (Peng et al., 2017; Shih et al., 2009; Yu et al., 2016).

Among Taiwan's civil service examinations, the air traffic controller category in the special examination for civil aviation personnel has included an aptitude test as part of the physical checkup process. This aptitude test assesses candidate's judgement, cognitive and reasoning abilities, and spatial awareness (Ministry of Examination[MOEX], 2020). There are also suggestions that aptitude and personality tests be gradually included in examinations like the special examination for consular and diplomatic personnel and the examination for police officers (Lee, 2012; Wu, 2012; Yu, 2015).

In the private sector, examples of aptitude tests being used for human resource selection and appointments can also be found (Tuan, 2006). For example, examinations held by Securities & Futures Institute (SFI) to recruit employees for the Taiwan Stock Exchange Corporation and the Taiwan Depository & Clearing Corporation include career tests, comprising aptitude and personality tests. Another example is Taipei Metro, which assesses mathematical and logical reasoning in the first stage examination and multiple latent traits and personal qualities in the second stage examination with an oral examination or psychometric test.

In summary, the evaluation of multiple latent traits has become a primary consideration for recruitment and selection in various fields. First, evaluation has evolved from a single dimension to multiple dimensions. Scoring systems have shifted from an aggregate score to a diagram covering a wide range of skill areas. Second, the focus of tests has changed from learning outcomes to potential for future success. Third, the weight of paper-and-pencil tests has been reduced with more non written tests being used, such as practical tests, oral examinations, and computer-based testing. These trends provide significant reference for adopting aptitude tests in civil service examination.

II. Feasibility Evaluation of Using Aptitude Testing in Civil Service Examinations

A. Legal and Evaluation Research Analysis

The idea of incorporating multiple latent trait assessment into public sector recruitment has a long history. When the Examination Affairs Act was established in 1988, non-written assessment methods such as publication and invention reviews, academic paper reviews, oral examinations, and practical examinations were included in Article 19. However, these alternative assessment methods were rarely implemented. In the 2015 amendments to the Act, Article 14 stipulated that assessments may take the form of written examinations, oral examinations, psychometric testing, physical fitness tests, practical tests, and publication and invention review. Details about the implementation of these alternatives are provided in other MOEX regulations.

Apart from legal adjustments, research has been conducted on the inclusion of different testing methods in public employee selection and assessment. In 2016, for example, the Civil Service Development Institute commissioned research on the use of psychometric testing in the public sector and related legal issues (Yu et al., 2016). In 2017, MOEX conducted a feasibility study on the use of psychometric testing in civil service examinations (Peng et al., 2017). Both studies suggested that psychometric testing, including aptitude tests, gradually be implemented for the selection and placement of some civil service positions. In 2019, MOEX commissioned experts in psychometric testing to develop a sample aptitude test for civil service examinations and conduct evaluation and simulation studies on the use of examination scores (Chen et al., 2020a). It is a pity, however, that aptitude tests are used only in the air traffic controller category in the special examination for civil aviation personnel and the promotional examination for diplomatic personnel. The following summarizes key points of the laws and evaluation studies.

(A) Laws and Regulations Evolvement

Legal regulations are the most important basis for public employees' lawful administration. The first step towards the inclusion of aptitude testing in the selection process is to amend laws. As mentioned above, in 2015, MOEX amended the Examination Affairs Act to include psychometric testing as an examination method, while Article 10 of the Civil Service Examinations Act added psychometric testing as an examination method. As for how psychometric testing should be administered, MOEX announced the Psychometric Examination Regulations on December 14, 2015, Article 3 of which states that "All examinations may, if needed, adopt psychometric

testing in the form of words, numbers, symbols, shapes, or maneuver to evaluate candidate's mental qualities, such as intelligence, aptitude, personality, attitude, and interests. Types of psychometric testing include: (1) intelligence test; (2) aptitude test; (3) personality test; (4) interest test; and (5) other forms of psychometric testing. One or more types of psychometric testing may be used depending on the nature of the examination." There is already a legal basis for incorporating aptitude tests into civil service examinations. From the above quoted text, it can be seen that the existing examination regulations distinguish achievement tests from psychometric testing, which differs from the psychological view of achievement tests as a form of psychometric testing (Anastasi & Urbina, 1997).

As stated in Article 4 of the Psychometric Examination Regulations, the type, scoring, criteria, test weighting, and other relevant matters should be subject to specific examination regulations when psychometric testing is used. That is to say, as long as the scoring standards and use of scores are defined and announced before the examination, the administration of aptitude tests is permitted by law. Hence, Ministers without Portfolio have suggested to include psychometric testing in civil service examinations. For example, Chen et al. (2011) suggested incorporating psychometric testing, which can help predict future performance, into the civil service examination to select those who can deliver expected performance. A civil service human resource performance improvement plan, which was passed on August 13, 2015, also added psychometric testing into its agenda of "improving psychometric testing and flexible examination system" followed by relevant discussions.

(B) Feasibility Study

Yu et al. (2016) was commissioned by the Civil Service Development Institute to investigate the use and the legal background of psychometric testing in the public sector. The aim of this study was to analyze the practice related legal regulations, and effectiveness of psychometric testing in the public sector in other countries. After a review of the practice and legal regulations in advanced countries in Europe, North America, and Asia, this study proposed a self-exploration mode, job assignment mode, and multi-stage composite functionality mode. Focus group discussions were conducted. The study also conducted a large-scale questionnaire survey on the results, timing, institutionalization, and information disclosure related to the implementation of psychometric testing. The survey found that respondents and scholars reported a high level of acceptance of government agencies using psychometric testing, though issues remained due to lack of accommodating measures. Suggestions were made that psychometric testing be used first for low-risk purposes, quality of evaluation methods

be publicized, and psychometric testing be implemented in a staged, gradual manner.

In a 2017 MOEX commissioned project, Peng et al. (2017) conducted a feasibility study on the use of psychometric testing in civil service examinations. The study compiled and analyzed practices in Taiwan and abroad to explore psychometric testing types, purposes of implementing psychometric testing, implementation, and how accommodating measures were designed (such as the need for information disclosure to ensure public credibility, the quality requirements of psychometric testing, and the maintenance of the examination mechanism, etc.). The study approached the topic through literature analysis, focus group discussions, and in-depth interviews. Suggestions were made that civil service examinations should use a two-step screening process, use tests that can measure high order cognitive abilities (like aptitude tests that can evaluate logic thinking and data analyzing skills), and use job analysis as selection criteria. It also suggested that existing achievement tests containing multiple-choice and essay questions should be designed to assess high order cognitive skills or competence to avoid a focus on rote knowledge and understanding. The study also suggested establishing a task force to promote the use of psychometric testing in civil service examinations.

(C) Study on Test Contents and Samples

In an aptitude test development project commissioned by MOEX to Chen et al. (2020a), researchers developed dimensions for aptitude tests applicable to civil service examinations. They also developed test samples, conducted reliability and validity evaluations, and made suggestions for establishing norms and using examination scores. The study navigated literature on the public servant core competences in Taiwan and other countries (Chen, 2013, 2016; Chiang, 2002; Cofsky, 1993; Kao, 2013; McClelland, 1973; MOEX, 2013; Shih, 2010; Spencer & Spencer, 1993) and categorized the core competences into professional skills, personality, and aptitudes. Professional skills relate more closely to achievement tests which evaluate candidates' learning outcomes in a specific subject. Professional skills are closely associated with an individual's learning environment and hard work. Personality as tested with personality tests reflects an individual's habitual response patterns, with no right or wrong answers. Aptitudes, including communication skills, judgement and thinking, innovation, mathematical and statistical skills, language comprehension, problem solving, and computational thinking are shown in Table 1. The study further combined these core competences with important dimensions in existing academic and vocational aptitude tests to develop aptitude test dimensions and test samples for civil servants, as shown in Table 2 and Figure 1 (Chen et al., 2020b).

Table 1

Categorization of Aptitudes Related to Civil Servant Core Competences

- (1) Communication skills: Media communication, leadership/coaching others, communication and coordination/contacting people, verbal expression/debate/negotiation, marketing and customer service, public relations, message delivery, negotiation skills, etc.
- (2) Judgement and thinking: Thinking/judgment/decision-making skills, policy analysis, environment change and trend analysis, logical deduction/analysis, cost/benefit analysis, assessment/evaluation, critical reflection, observation skills, etc.
- (3) Innovation: Creative thinking, innovative thinking, creativity, innovative improvement, innovation, etc.
- (4) Mathematical and statistical skills: Mathematical reasoning, statistics, etc.
- (5) Language comprehension: Word processing, business writing/official document writing, comprehension/reading comprehension skills, etc.
- (6) Problem solving: Strategic thinking and planning, strategy execution, problem solving, adaptability, solution design, vision planning, contingency/crisis management, complex problem analysis and solving, investigation/research, etc.
- (7) Computational thinking: Information application, computer skills, information technology, etc.

Note. From “The Development and Application of the Aptitude Test for Civil Servants.” by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020b, *National Elite Quarterly*, 13(2), 144.

Table 2

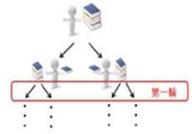
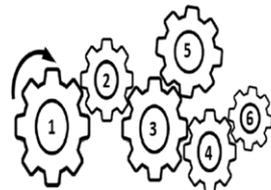
Dimensions of Civil Service Aptitude Tests

Dimension	Sub-dimension	Details
Mathematical logic/ Computational thinking	Mathematical logic	Understanding and reasoning of numerical relationships, ability to handle quantitative materials
	Computational thinking	Ability to solve problems using computational and thinking skills, such as abstraction, algorithm design, and pattern recognition
Spatial intelligence/ Mechanical reasoning	Spatial intelligence	Ability to mentally visualize and manipulate shapes or three-dimensional spaces, visualize the appearance and position of objects based on diagrams or drawings
	Mechanical reasoning	Ability to reason using basic mechanical principles, tools, and physics
Problem solving / critical thinking	Problem solving	Ability to clarify key points, identify useful information, propose solutions, and evaluate the pros and cons of different approaches when faced with problems
	Critical thinking	Critical thinking skills to identify evidence, hypotheses, induce, deduce, explain, and evaluate
Verbal reasoning /Communication skills	Verbal reasoning	Ability to understand semantic meanings and relationships between words, conduct sampling or inferential reasoning
	Communication skills	Ability to comprehend messages accurately and articulate opinions using appropriate expressions and communication skills
	Creative thinking	Ability to generate innovative ideas or thinking from different perspectives

Note. From “The Development and Application of the Aptitude Test for Civil Servants.” by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020b, *National Elite Quarterly*, 13(2), 145.

Figure 1

Samples of Aptitude Test for Civil Service Examinations

<p>試題設計-數理邏輯</p> <p>✓ 請你依照下方九九格直欄及橫列的數字規則，判斷出？為何數字。</p> <table border="1" style="margin-left: 20px;"> <tr><td>4</td><td>6</td><td>2</td></tr> <tr><td>7</td><td>9</td><td>5</td></tr> <tr><td>15</td><td>17</td><td>?</td></tr> </table> <p>(A)12 (B)8 (C)19 (D)13</p> <p>✓ 行政院農委會某秘書參加了台北 101 垂直馬拉松競賽，秘書從 1 樓走到 8 樓需要花 56 秒的時間，請問若以同樣的速度往上走樓梯，秘書還需要幾秒的時間才能走到第 20 層樓？</p> <p>(A)84 (B)91 (C)96 (D)104</p> 	4	6	2	7	9	5	15	17	?	<p>試題設計-運算思維</p> <p>✓ 國民健康署某科長想發會議資料給不同部門的每一位同事，因每一份會議資料皆印有同事的名字，因此只能專人傳送。為了節省時間，他將會議資料分成兩大疊，並分別送到該疊會議資料第一份名字的同事手中(如下圖)，再請這兩位同事收下自己的會議資料後，將剩餘的會議資料平均分成兩份(若無法平分，兩疊的數量最多差一張)，分別送到該疊會議資料的第一位名字同事的手中(第一輪)，以此類推。而若會議資料總共有 40 份，請問最多需要經過幾輪的傳送，才可以让每個人拿到自己的會議資料？</p> <p>(A)2 (B)3 (C)4 (D)5</p> 
4	6	2								
7	9	5								
15	17	?								
<p>試題設計-空間概念</p> <p>✓ 公共工程甲委員前往新竹高鐵站進行相關工程勘測，他先在高鐵站外拍了一張照片，請你依據照片及衛星圖的相對位置，判斷出甲委員在何處拍照的。</p> <p>(A) A 處 (B) B 處 (C) C 處 (D) D 處</p> 	<p>試題設計-機械推理</p> <p>✓ 交通部甲科員正在進行船舶的維修工程，若齒輪 1 為順時針方向轉動，請問下列齒輪中何者方向與齒輪 1 的轉動方向相同？</p> <p>(A) 齒輪 2 (B) 齒輪 4 (C) 齒輪 5 (D) 齒輪 6</p> 									
<p>試題設計-問題解決與批判思考</p> <p>✓ 大明是話劇社團社長，最近在籌畫社區成果展由六位大四畢業生排演話劇，就在距離成果展前，阿強臨時告知大明因為家中父親意外過世，無法參加社區成果展，請他另作安排，大明如果希望社區成果展能夠順利演出的話，哪些不是他現在應該考慮的問題？</p> <p>(A)阿強演的戲份對整齣戲的影響 (B)阿強是否會苦惱有辦法參加演出 (C)由較熟悉劇本的大三生頂替阿強演出的可行性 (D)阿強角色對整齣戲的重要性</p> <p>✓ 行政院農業委員會農作物轉種指導時，在農田裡發現了一大塊礦石，此時甲科員說：「這不是鐵，也不是銅」；乙科員說：「這不是鐵，是錫」；丙科員說：「這不是錫，是鐵」；此時隨行的地理老師則說：「你們三人有一人全對，一人全錯，一人半對半錯」，請你根據上述的資訊，判斷出此礦石為何？</p> <p>(A)鐵 (B)錫 (C)銅 (D)其他礦物</p>	<p>試題設計-語文推理與溝通表達</p> <p>✓ 「歌舞劇：演員」兩者間的關係與下列哪個選項的關係配對相同？</p> <p>(A)球場：運動選手 (B)演唱會：歌手 (C)電腦：打字員 (D)辦公室：職員</p> <p>✓ 阿強與阿達今天約好帶家人一起去爬山，兩家人預計在山腳下會合，但因為阿強家中小孩臨時生病無法赴約，不得已得跟阿達取消原本的爬山計畫，如果阿強希望不要破壞雙方友情，且站在彼此的角度進行思考的話，他該如何跟阿達表達比較好呢？</p> <p>(A)「我家小孩昨天生病，今天體力還沒恢復，可能爬不上山頂。」 (B)「我家小孩昨晚臨時發高燒，身體不適，怕傳染給你小孩。」 (C)「我家裡有事情，臨時無法出席，你帶著你的家人去吧。」 (D)「我家今天無法出席，跟你改約下周再去爬山，好不好？」</p>									
<p>試題設計-創造思維</p> <p>✓ 請由題目中所提供的三個詞彙進行發想，試著從下列選項中找到一個與題目中三個詞彙各自有相關的字詞。(此處的相關可以是語意關聯、或可以組成成語或用語)</p> <p>開始、太陽、鬧鐘</p> <p>(A)黃昏 (B)早晨 (C)公園 (D)手錶</p> <p>✓ 甲科員在中午休息時間用 10 根火柴棒排成了一张椅子，請問甲科員最少移動幾根火柴棒的位置，便可使這張椅子變成顛倒的呢？</p> <p>(A)2 根 (B)3 根 (C)4 根 (D)5 根</p> 										

Note. From “The Development and Application of the Aptitude Test for Civil Servants.” by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020b, *National Elite Quarterly*, 13(2), 148.

B. Practical Considerations When Incorporating Aptitude Tests into Civil Service Examinations

To understand what potential challenges may arise during the implementation of aptitude tests in civil service examinations, this paper referred to the test implementation plan structure proposed by Chen (2019) and combined research findings from the previous section. This paper provides reference for policy making and planning in consideration of factors to such as human, administrative, time, space, and tools if aptitude tests are to be incorporated into civil service examinations.

(A) Human Factors

1. Drafter Training

As with civil service examination achievement tests, implementing aptitude tests requires the recruitment of drafters. It would be better if these drafters possess the expertise related to aptitudes in Table 2, or are psychologists or educational scholars familiar with aptitude test dimensions. After drafters have been appointed and sign a non-disclosure agreement, they will receive training based on test dimensions and test samples proposed by Chen et al. (2020a) to develop drafting guidelines, generate the required number of tests regularly each year, and assist with test revision and review.

2. Examinee Awareness and Release of Past Examination Questions for Practice

Research findings by Peng et al. (2017) and Yu et al. (2016) indicate that releasing information regarding aptitude test contents and quality to ensure test credibility is essential to implementing the testing in civil service examinations. Article 4 of the Psychometric Examination Regulations, published by MOEX in 2015, states that the type, scoring, criteria, test weighting, and other relevant matters should be subject to specific examination regulations when psychometric testing is used. Test contents can refer to test dimensions and sample tests proposed by Chen et al. (2020b). Past examination questions may be released for candidates to practice, and to help reduce their test anxiety. The calculation and use of test scores should be in place before the test is implemented. Information packs should also be produced and announced online at MOEX's website before official implementation.

Releasing test questions can help examinees familiarize themselves with questions types and reduce their examination anxiety. These test questions can be used as pilot tests to collect examinee responses and question quality information such as difficulty and discrimination levels. As long as the pilot tests are conducted in a MOEX-controlled environment (computer room for computer-based testing was recommended) with confidentiality and memory interference in place, test questions may be released regularly as a means for MOEX to manage test quality and maintain difficulty levels and consistency.

(B) Administrative Factors

Administrative factors can be considered at three time points: before, during, and after the test. With MOEX's rich experience in executing civil service examinations, using aptitude testing in civil service examinations should be workable in practice. Question and answer papers should be ready before the test. If the test is computer-based, the aptitude test should be uploaded to the test system. Training sessions should be provided to invigilators on how to ensure the smooth running of the aptitude testing, respond to candidate questions or contingencies, and maintain records during testing. These procedures are similar to those of other tests and will not be further elaborated upon in this article.

Test question confidentiality, scoring methods, and use of test scores are matters that require additional planning for the implementation of aptitude testing. The following provides more detail:

1. Test Question Confidentiality

Unlike achievement test questions which can be generated quickly based on a wide range of learning content, aptitude test question design requires considering external behaviors that reveal an individual's abstract latent traits. Some are even designed with real-life workplace scenarios to allow examinees to demonstrate behaviors relevant to core competences. Pilot tests must also be held to determine the question difficulty, discrimination, and to analyze test reliability and validity to ensure the effectiveness of the test. Therefore, it is not appropriate to release all questions immediately after the test is completed. If existing regulations require the release of aptitude test questions after the test is completed, amendments should be made or computer-based testing may be used because questions for computer-based testing can lawfully remain confidential.

Even though regulations allow aptitude test questions to remain confidential after testing, the confidentiality of aptitude test questions is more important than that of achievement tests. This is because these questions have standardized answers and will be used repeatedly in other examination categories, making it a necessity to maintain the confidentiality from the beginning. Invigilators need to ensure examinees do not copy or memorize the questions during the examination and do not take any papers when leaving the test. If necessary, standardized procedures may be implemented to interfere with memory.

2. Scoring and Interpretation

Tests can be divided into norm-referenced and criterion-referenced tests depending on how test results are scored and interpreted. The former compares an individual's test results against other test takers to see the individual's standing in the group. The latter compares an individual's performance against pre-determined standards or a cut-off score to see if the test taker possesses the required skills.

Considering the characteristics of these two approaches, norm-referenced tests may be more suitable for civil service examinations while criterion-referenced tests may be more suitable for professional and technical personnel examinations.

What sets aptitude testing apart from other norm-referenced achievement tests is that the norm of aptitude tests refers to all test takers, not test takers of a specific examination category like general subjects such as Chinese or English. In scoring and interpreting scores, it is important to consider the characteristics of norm-referenced tests and the reference group. Screening standards and weighting methods must also consider the hiring needs of each examination category.

In addition to classical test theory, the item response theory (IRT) can also be used to convert raw scores into norm-referenced scores. Since civil service examinations are held several times each year, it may be necessary to generate different aptitude tests for different examination categories under the premise of test security. Hence, issues of test equating and norm availability must be considered. The IRT model can estimate test takers' ability and question difficulty level based on the test taker's response to the questions. The IRT method has characteristic invariances in ability estimation and test question parameters and can be used to compare test results which measure the same abilities using different tests on the same scale. So IRT is well-suited for aptitude testing which is held multiple times each year measuring same abilities through different examination questions. IRT can be used in this situation to equate tests and convert scores.

3. Test Score Use

In a study by Yu et al. (2016), psychometric test scores were used for different purposes of promotion, job assignments, and training for current civil servants. This section on the use of psychometric testing scores is based on the research findings of Peng et al. (2017) and Chen et al. (2020a). According to suggestions made in the study by Peng et al. (2017), it is appropriate to implement psychometric testing in the form of staged examinations and trials should be conducted before official launch to ensure that test scores can predict future performance.

A meeting was held by Chen et al. (2020a) with scholars who assisted MOEX with examination affairs or conducted civil service recruitment studies to discuss the use of aptitude test scores. During the meeting it was suggested to use threshold and weighting methods to calculate aptitude test scores. In the threshold method, the selection process is divided into two stages. The first stage considers the scores of examination subjects tested in the existing civil service examinations and the second stage considers aptitude test scores. To qualify, candidates must pass the first stage with high scores and pass the cut-off score at the second stage. The weighting method, on the other hand, incorporates aptitude test scores and weighted scores of other examination subjects tested to determine if the candidate qualifies.

To see how different weights or cut off scores would impact examination results, Chen et al. (2020a) conducted a simulation study. Based on the correlation coefficients between achievement test dimensions and aptitude test dimensions from 434 research subjects, this study stimulated 1,000 examination takers' data in a multivariate normal distribution with eight dimensions representing their performance in Chinese, English, professional subjects, problem solving/critical thinking, verbal reasoning/communication skills, mathematical logic/computational thinking, spatial intelligence/mechanical reasoning, and creative thinking. The correlation between achievement and aptitude test scores was set to low (0.1~0.3) and moderate (0.4~0.6) to simulate the impact of different correlation levels on aptitude test scores. The pass rates for both methods were set at 10%. The results are as follows:

(1) The Threshold Method

Since the aptitude test covered five dimensions, the setting of each in the study was divided into non-compensatory (each dimension score must pass certain sets of standards) and compensatory (the average of the scores of all dimensions passes a certain sets of standards, allowing the dimensions to complement each other). The results show that when the aptitude test threshold is at negative one standard deviation (-1.0 SD, PR value ≥ 17), the impact ratio of qualification results would reach 26% only when the correlation between the achievement test and the aptitude test was small and the aptitude test threshold was non-compensatory. In all other cases, the impact ratio was below 5%, as shown in Table 3:

Table 3

Impact of Aptitude Test Threshold at Negative One Standard Deviation (-1.0 SD) on Qualification Results

	Non-compensatory threshold	Aptitude test: passed	Aptitude test: failed	Impact ratio
Low degree correlation	Achievement test: passed	74	26	26%
	Compensatory threshold	Aptitude test: passed	Aptitude test: failed	Impact ratio
	Achievement test: passed	99	1	1%
Moderate degree correlation	Non-compensatory threshold	Aptitude test: passed	Aptitude test: failed	Impact ratio
	Achievement test: passed	96	4	4%
	Compensatory threshold	Aptitude test: passed	Aptitude test: failed	Impact ratio
	Achievement test: passed	100	0	0%

NOTE. Adapted from *Final Report on the Development of Aptitude Tests for Civil Service Examinations* (pp.87-88), by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020a, Ministry of Examination.

(2) The Weighting Method

Table 4 shows the impact of aptitude test weighting at 10%, 20%, and 30% on candidates are qualified using only achievement test scores. At a moderate degree correlation between achievement and aptitude tests, when aptitude test weighting is 10%, six out of the 100 candidates who passed using only the achievement test scores will fail; when the weighting is 20%, the impact ratio is 13%; when the weighting is 30%, the impact ratio will increase to 19%. At a small correlation, the qualification results rise to between 13% and 36%. All in all, the higher the aptitude test weighting, the greater the impact on qualification results. The impact is even greater when the correlation between achievement test and aptitude test is lower.

Table 4
Impact of Different Aptitude Test Weighting on Qualification Results

Moderate degree correlation	Aptitude test weighting: 10%	With weighting: passed	With weighting: failed	Impact ratio
	Achievement test: passed	94	6	6%
	Aptitude test weighting: 20%	With weighting: passed	With weighting: failed	Impact ratio
	Achievement test: passed	87	13	13%
	Aptitude test weighting: 30%	With weighting: passed	With weighting: failed	Impact ratio
	Achievement test: passed	81	19	19%
Low degree correlation	Aptitude test weighting: 10%	With weighting: passed	With weighting: failed	Impact ratio
	Achievement test: passed	87	13	13%
	Aptitude test weighting: 20%	With weighting: passed	With weighting: failed	Impact ratio
	Achievement test: passed	73	27	27%
	Aptitude test weighting: 30%	With weighting: passed	With weighting: failed	Impact ratio
	Achievement test: passed	64	36	36%

NOTE. Adapted from *Final Report on the Development of Aptitude Tests for Civil Service Examinations* (pp.81-82), by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020a, Ministry of Examination.

According to the suggestions of the study, if the threshold method is to be used to determine the minimum threshold for aptitude tests which are held in the second stage of the civil service examination and the correlation between aptitude and achievement tests is small, it is best to set a compensatory threshold with a standard deviation of negative point five (PR value ≥ 32) for the aptitude test. This way, the impact ratio will not exceed 10%. If the correlation between aptitude and achievement tests are medium, a compensatory threshold with a standard deviation of negative one (PR value ≥ 17) will result in an impact ratio of less than 5%. If a non-compensatory threshold is set, which means candidates must pass all dimensions, it is preferable to control the threshold under negative point five standard deviations (PR value ≥ 32) to keep the impact ratio under 15%. If the weighting method is used to incorporate the aptitude test scores into the final scores, it is recommended to assign a weight of no more than 10% to the aptitude test when there is a low correlation between the aptitude and achievement tests. When there is a moderate correlation, it is preferable to keep the weight of the aptitude test below 20%. In both cases, the impact ratio will not exceed 15%.

In general, when deciding the use of aptitude test scores, a pilot study may be conducted to collect aptitude and achievement test scores from examination takers of a specific year and analyze the data using actual correlation coefficients with different weightings and threshold values to ensure rationality and accuracy of the policy. Regarding whether the use of scores should differ depending on the examination category or whether a uniform threshold or weighting method should be used, it is best to first collect data from representative samples to see if the aptitude test can really predict future performance before making the final decisions.

(C) Timing

Regarding the timing aptitude tests, Peng et al. (2017) suggested that aptitude tests should not be tested during the existing achievement test stage; rather, they should be computer-based second stage tests. Questions should not be disclosed after the test. Chen et al. (2020a) also made suggestions at the meeting on the use of aptitude test scores with experts that threshold method should be used for aptitude tests in the second stage of civil service examinations. Even using the weighting method, as it is not recommended to publicize aptitude test questions, it is preferable that to have the aptitude test take place in a different stage than the achievement test.

Aptitude test length should be subject to the number of dimensions and questions and the difficulty level of the test. According to research by Chen et al.(2020a), when an aptitude test covered five dimensions in 55 questions, if the examination session is 70 minutes, 90% of research subjects could complete the test. In the study, questions in

the verbal reasoning/communication skills section were easier (average correct response rate 0.80) and questions in the creative thinking section were harder (average correct response rate 0.47). How long an aptitude test will take should be determined based on the number of dimensions and questions.

(D) Venue and Equipment

The venue and environment for administering aptitude tests are determined based on the test format. If the test is administered as a written test, venue requirements are the same as existing achievement tests. However, to ensure test confidentiality, computer-based testing would be used and so the current capacity of available computer-based testing venues must be considered. According to MOEX statistics, in 2011, 510,000 candidates applied to take civil service examinations and in 2021, the number dropped to 240,000 (Jiang, 2022). Even if computer-based aptitude tests are held twice a year, with up to four sessions each day, a capacity of 30,000 for each session is required.

With so many candidates taking computer-based tests, venue capacity is not the only thing to consider, computer equipment stability, data transmission speed, and security are also important issues to deal with. To ensure stability of computer equipment, regular inspection and maintenance and pre-test stress testing should be conducted. Contingency measures should also be put in place for situations such as computer unit malfunctioning or a complete black out at the examination venue. In terms of data transmission and security, servers can be set up at each venue using local intranet to share server load of client computers for enhanced speed and security. Since MOEX has rich experience in conducting computer-based testing for large numbers of test takers and the capacity for computer-based testing will exceed 10,000 in 2023, implementing computer-based aptitude tests should not be a problem.

(E) Tool Factors

Currently, there are no official aptitude tests for civil service examinations. Only two sets of aptitude test question samples and demonstration tests exist, developed by Chen et al. (2020a) for a MOEX-commissioned research project. If aptitude tests are to be used in civil service examinations, it may be necessary to evaluate whether to use existing aptitude tests available on the market or develop new aptitude tests for the civil service's needs. Other matters related to test questions and norm renewals should also be planned ahead of time.

1. Using Existing Tests or Developing New Tests

If existing aptitude tests are used, those most commonly used in Taiwan are the

academic aptitude and vocational aptitude tests. The dimensions of academic aptitude tests are mostly related to learning activities at school, while vocational aptitude tests are related to career choices or occupational activities. Early aptitude tests included the multi-factor aptitude test (Lu et al., 1994) and the general aptitude test (Vocational Training Bureau, Ministry of the Interior, 1985). The former was targeted at middle school students with dimensions covering verbal reasoning, mathematical reasoning, abstract reasoning, mechanical reasoning, spatial relations, grammar and rhetoric, and perceptual speed and accuracy. These tests are designed to help students make informed academic or vocational decisions. Validity analysis of these tests focuses on the students' learning outcomes in Chinese, English, and math. The latter was developed by the U.S. Department of Labor and translated and adapted for use in Taiwan by the Ministry of the Interior's Vocational Training Bureau. Its scores cover nine areas, including intelligence, language, numeracy, spatial awareness, graphic perception, clerical perception, movement dexterity, and manual dexterity. Validity analysis was conducted using the learning performance of vocational high school students in common subjects or occupational subjects.

Recent aptitude tests include the revised multiple aptitude battery (Wu et al., 2011) and the computerized adaptive career aptitude test (Sung, 2015). The former covers dimensions of verbal reasoning, numerical reasoning, graphic reasoning, mechanical reasoning, spatial relations, Chinese vocabulary, English vocabulary, perceptual speed and accuracy. Although it can be used to provide reference for student counseling or enterprise recruitment purposes, the dimensions tested focus on school learning scenarios. Validity analysis was also based on students' admission examination scores. The computerized adaptive career aptitude test focuses on vocational dimensions including language, math, spatial ability, logical reasoning, observation, scientific reasoning, aesthetics, and creativity. The test targets junior high school students and the results can help students understand their strengths and help them choose future educational paths. Validity analysis was also based on students' admission examination scores.

The focus and function of available aptitude tests in Taiwan all serve students' needs to make career decisions. They may not be suitable for public employee selection. It may be necessary to develop aptitude tests that cater to civil service hiring needs. If the government is to develop aptitude test that suit civil service hiring needs, it can seek reference from the research project by Chen et al. (2020a). The research project considered the five common dimensions in old aptitude tests and the results of civil servant core competency analysis to formulate a five-dimension aptitude test and sample questions, shown in Table 2 and Figure 1. To reflect real-life work scenarios, the study developed two sets of sample multiple-choice questions: general and scenario-

based. Scenario-based multiple-choice questions consider the actual scenarios public servants may encounter and the abilities to be tested, shown on the right side of Figure 2. Scenario-based questions cannot be too specific, or they cannot generalize to other similar scenarios. They cannot be too general either, or they will fail to evaluate the needed abilities and cannot assess if the candidate can apply these abilities in real life situations.

Figure 2

Multiple-Choice Sample Question for the Civil Service Aptitude Test

<p>試題設計-語文推理與溝通表達</p> <p>✓ 「歌舞劇：演員」兩者間的關係與下列哪個選項的關係配對相同？</p> <p>(A) 操場：運動選手</p> <p>(B) 演唱會：歌手</p> <p>(C) 電腦：打字員</p> <p>(D) 辦公室：職員</p>	<p>試題設計-空間概念</p> <p>✓ 公共工程甲委員前往新竹高鐵站進行相關工程場勘，他先在高鐵站外拍了一張照片，請你依據照片及衛星圖的相對位置，判斷出甲委員在何處拍照的。</p> <p>(A) A 處</p> <p>(B) B 處</p> <p>(C) C 處</p> <p>(D) D 處</p> 
<p>General Multiple-Choice Questions</p>	<p>Scenario-Based Multiple-Choice Questions</p>

NOTE. From *Final Report on the Development of Aptitude Tests for Civil Service Examinations* (p.21), by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020a, Ministry of Examination.

While two sets of sample questions have been developed by Chen et al. (2020a), there is still room for improvement in difficulty and test reliability levels of existing sample questions. Both classical test theory and item response theory analyses have shown that the questions in the verbal reasoning/communication skills section are relatively easier, with a higher correct response rate. This is probably because these skills are basic skills in daily life. Questions in the creative thinking section, on the other hand, are relatively harder. It could be that these questions involve thinking activities that are less common in daily life. Civil servants or college students have long been required to develop correct concepts and choose the right answers, or to follow established rules when fulfilling their duties. Therefore, they may find it more challenging to work on questions that require thinking from different perspectives.

In terms of reliability, the research used internal consistency reliability and the separation reliability converted from test information as the indicator of measuring accuracy. Since the civil service aptitude tests have questions that adopts polytomous scoring, the research used Cronbach's α coefficient to conduct reliability analysis for question sets A and B. The creative thinking dimension had lower reliability because of its multi dimension nature. As to other dimensions, with just 8 to 14 questions, the reliability falls between 0.4 and 0.75. Since both question sets in this research shared some questions, concurrent calibration could be used to combine questions and equate

research subjects' abilities. Then, based on the variance and estimate error, differential reliability of the two sets combined was estimated. The results showed that except for the creative thinking dimension, the other four dimensions, under the condition of having 12 to 23 questions, had research subject differential reliabilities ranging from 0.67 to 0.72 and a test question reliability of higher than 0.95 (Chen et al., 2020a).

In terms of validity, the research presented differential validity by comparing different groups. Aptitude test results of different civil servant groups showed that civil servants received slightly higher scores than non-civil servants and college students in all dimensions. The difference between civil servants and non-civil servants was especially pronounced, as shown in Table 5.

Table 5

Descriptive Statistics of Aptitude Test Results of Different Groups

Profession	No.	Dimension				
		Problem solving/critical thinking	Verbal reasoning/Communication skills	Mathematical logic/Computational thinking	Spatial intelligence/Mechanical reasoning	Creative thinking
		Average (standard deviation)	Average (standard deviation)	Average (standard deviation)	Average (standard deviation)	Average (standard deviation)
Non civil servants	92	97.7 (6.8)	97.6 (7.0)	97.2 (10.1)	92.9 (17.8)	98.1 (4.9)
Civil servants	235	100.9 (6.3)	101.0 (7.0)	100.9 (10.1)	103.0 (13.4)	100.9 (3.8)
Students	107	100.2 (7.0)	100.1 (7.6)	100.7 (10.5)	100.4 (16.6)	100.0 (4.4)

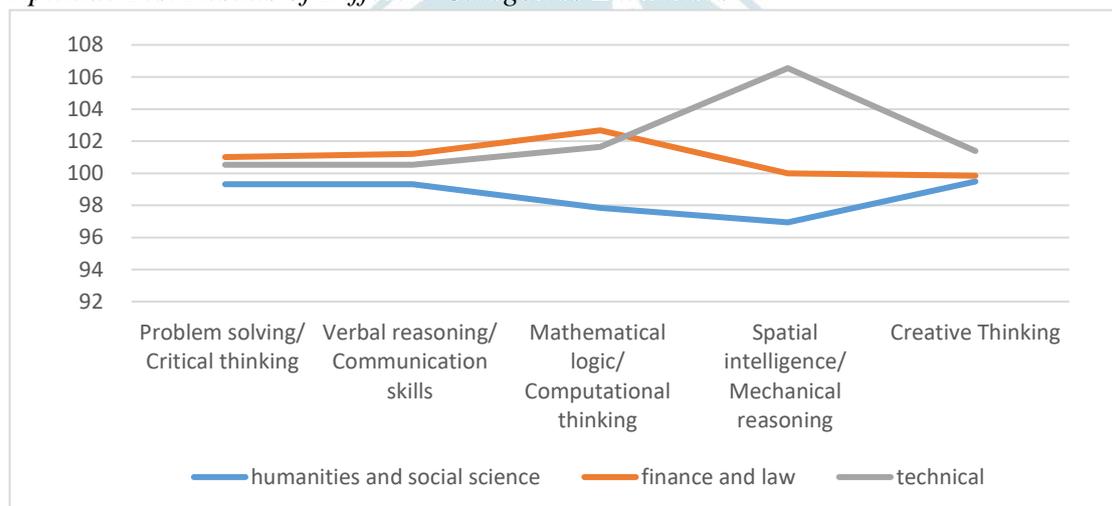
NOTE. Ability scale average 100, standard deviation 15. From *Final Report on the Development of Aptitude Tests for Civil Service Examinations* (p.64), by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020a, Ministry of Examination.

Chen et al. (2020a) also found that examinees taking different examination categories performed differently in the dimensions. As shown in Figure 3, in “mathematical logic/computational thinking”, examinees in finance and law and technical categories performed better than those in humanities and social science categories; in “spatial intelligence/mechanical reasoning”, examinees in technical categories performed the best and those in humanities and social science categories performed weaker; in “problem solving/critical thinking” and “verbal reasoning /communication skills”, no major difference was observed; in “creative thinking”, examinees in technical categories performed slightly higher than those in humanities and social science categories and those in finance and law. Test results in Figure 3 show

that while humanities and social science categories and finance and law categories are of administrative nature, civil servants from these two perform differently in the dimensions. This is in agreement with the argument proposed by scholars in the expert meeting in the research that aptitude tests should be considered separately because of the different core competences of the categories. In addition to the little difference in the performance of some dimensions between current civil servants in technical categories and non-civil servants, current civil servants in humanities and social science categories and in finance and law categories perform better than non-civil servants. This means that the test dimensions are discriminatory in selecting personnel in humanities and social science categories and finance and law categories. As for the selection of civil servants in technical categories, “spatial intelligence/mechanical reasoning” dimension is more discriminatory.

Figure 3

Aptitude Test Results of Different Categories Examinees

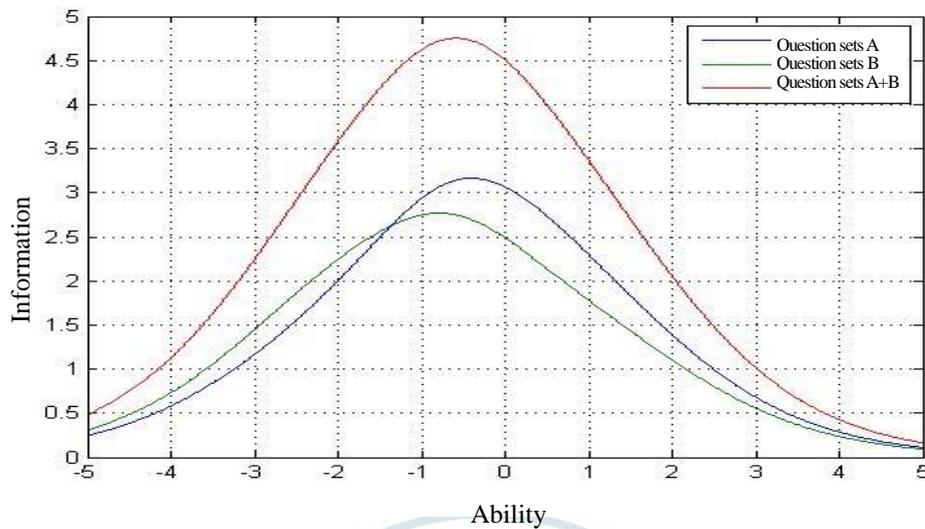


NOTE. Adapted from *Final Report on the Development of Aptitude Tests for Civil Service Examinations* (p.65), by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020a, Ministry of Examination.

If the aptitude test is to be used for civil servants, the question sets developed by Chen et al.(2020a) can be used as a foundation, with extended response time and increased number of questions to roughly 20 questions per dimension, to allow the reliability of internal consistency of each dimension to reach 0.80 or higher. The information graph of the test item response theory shows that most dimensions within the moderate ability range (-2.0 ~ +1.0) can provide higher levels of measuring precision (Figure 4 is the example of problem solving/critical thinking dimension). This, in general, aligns with the abovementioned second factor, where the threshold method using the criteria of 0.5 or 1 standard deviation below average as the preliminary screening standards.

Figure 4

The Test Information of Civil Service Aptitude (Problem Solving/Critical Thinking Dimension)



Note. Adapted from “The Development and Application of the Aptitude Test for Civil Servants.” by P.-H. Chen, H.-Y. Huang, S.-Z. Su and T.-Y. Ting, 2020b, *National Elite Quarterly*, 13(2), 154.

2. Question Bank Establishment and Norm Updating

The civil service aptitude test differs from general aptitude tests in that the civil service aptitude test scores can impact an individual’s employment prospects, so the risks associated with the test are high. Additionally, with the tests taking place on a large scale at different times of the year, it is necessary to have multiple sets of test questions that can be used alternately to ensure fairness and prevent test takers from memorizing the questions. A preferable way to construct the test bank like some examination subjects. To this end, it is suggested that the target number of questions in the question bank be based on the number of examinations held each year. For example, if the tests take place four times a year, the number of questions in the test bank should be at least four times the number of questions in the official question booklet. If the questions need to be kept confidential after the examination, questions may be appropriately overlapped for different examinations. Test item response theory analysis can be used to ensure psychometric validity in the equating of tests and scores of different question booklets.

Norms is another major consideration for the civil service aptitude test. Since the test is used to understand the relative position of candidates among all test-takers and within specific category group, it is necessary to establish separate norm-reference tables for the all examination takers and examination takers of specific categories. This will help to see test takers’ performances in multiple latent traits in different norm groups, improve score interpretation, and increase forecast validity. Some examination

categories may have few applicants, so the representativeness of the norms may be limited. Therefore, it is suggested to consolidate category examinations into broader groups (e.g., humanities and social science group, finance and law group, technical group) and establish norms based for them. When there's more data and more candidates applying for the examination and significant differences are observed in the dimensions, separate norms for that category can be considered.

III. Conclusion and Suggestions

Multiple latent trait is an important concept in the selection and adaptive development of civil service, especially in the face of a declining and ageing population, which has resulted in fewer people willing to choosing a career in the public sector. This trend can be seen from the sharp decline in the number of applicants for civil service examinations in the past ten years. With civil service workforce retiring each year and the demand for the number of civil servants remaining the same, compounded by a declining and ageing population and public and private sectors competing for talents, the shortage of manpower in the public sector can be expected. Hence, the evaluation of multiple latent traits and appropriate placement will be an important measure for the selection of public servants. Achievement tests, which are based solely on learning outcomes, will not be sufficient to meet the human resources needs of the public sector. Aptitude tests, on the other hand, can help both examination takers to know better about their own multiple latent traits and hiring agencies to place the right people in the right place.

To effectively implement aptitude tests in the selection of civil servants, careful planning and consideration is required in various aspects such as human, administrative, time, space, and tools. Based on the results of several relevant research reports in the past and the analysis of the current situation in the five dimensions in this article, if the government is to use aptitude tests as an assessment tool for selecting civil servants, the author suggests the following directions:

A. Implementation and Score Use Methods

According to the analysis in this article, it is best to schedule aptitude tests and general achievement tests for different dates and times. For example, aptitude tests can be conducted in the second stage in the form of computer-based testing. Question sets are randomly selected from the question bank and keep the overlapping of questions to a certain degree (e.g., 20%). Test questions will remain confidential after the test. Only the example questions or commonly used questions will be published for candidates to practice. This way, the confidentiality and validity of test results can be ensured.

Threshold method and weighting methods can be used to calculate scores. However, it is recommended to conduct trial tests for the first 1 or 2 years and not consider the scores to collect data regarding the correlation between aptitude tests and general achievement tests. After the appropriate threshold and weight of the aptitude tests for different categories and the predictability of future job performance is analyzed, the use of scores can be determined.

B. Establishing Question Banks and Norms

Currently, with no applicable and quality aptitude tests for civil service, it is necessary to establish a team to formulate test questions and develop norms for examination categories. Civil service aptitude test development projects commissioned by MOEX can be used as the foundation and the five dimensions can be used for the design of questions. Response data collected in the recent one or two years can be used to analyze test quality and effectiveness. Norms can be established. The selection effectiveness and predictability of norms of all examination takers and the norm reference table should be analyzed to ensure question sets meet the needs of the test and the explanatory power of the norms.

As societal changes accelerate, technological advancement and shifts in population age structure and nature of public service change in the recent 10 years, the core competencies required for civil servants gradually evolve as well. It is essential to appropriately adjust the question bank and norms of the civil service aptitude test for the needs of society and nation.

C. Establishing A Team For Promoting Psychometric Testing in Civil Service Examinations

Since the aforementioned work involves the collaboration and coordination among multiple divisions within MOEX, such as the Department of Examination Planning, Department of Senior and Junior Examinations, Department of Question Bank Management, and Department of Information Management, it may not be necessary to establish a separate legal entity or independent organization to promote the work. However, it is recommended to have the Minister or Deputy Minister serve as the convener and establish a team for the promotion of psychometric testing with external experts serving as consultants and heads of relevant divisions serving as members of the promotion team. Two to three contract employees can be hired to plan and execute the tasks. Regular meetings should be held to report on the progress, and coordination of manpower across divisions should be ensured. Initially, the development of the question bank, norms, and score use methods can be the focus of work. In the future, other psychological assessments, such as personality tests or value assessments, can be considered depending on the circumstances to optimize the effectiveness of the civil service selection process.

References

- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.
- Armed Services Vocational Aptitude Battery. (2023). *History of military Test*.
<https://www.officialasvab.com/researchers/history-of-military-testing/>
- Assessment & Teaching of 21st Century Skills. (2010). *21st century skills white paper*. <http://atc21s.org/index.php/resources/white-papers/>
- Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. In H. H. Goddard (Ed.), *Development of intelligence in children (the Binet-Simon Scale)*. Williams & Wilkins.
- Chen, J.-M., Huang, F.-Y., Suen, M.-W., & Lee, R.-J. (2011). Psychological test and national examination. *National Elite Quarterly*, 7(1), 15-31.
- Chen, P.-H. (2019). *Psychology and educational testing: Theories and practices in test construction*. Jing Tse Education.
- Chen, P.-H., Huang, H.-Y., Chen, Y.-H., Yeh, T.-T., Su, S.-Z., & Lee, P.-W. (2015). Evaluation of the psychometric properties of the general literacy test for university. *Journal of Research in Education Sciences*, 60(3), 95-125.
- Chen, P.-H., Huang, H.-Y., Su, S.-Z., & Ting, T.-Y. (2020a). *Final report on the development of aptitude tests for civil service examinations*. Ministry of Examination.
- Chen, P.-H., Huang, H.-Y., Su, S.-Z., & Ting, T.-Y. (2020b). The development and application of the aptitude test for civil servants. *National Elite Quarterly*, 13(2), 139-162.
- Chen, Y.-J. (2013). The inclusion of competency analysis in national examination oral exams. *National Examination Forum Quarterly*, 3(4), 16-28.
- Chen, Y.-J. (2016). The consolidation of examination categories and examination subjects from occupational competency perspectives. *National Examination Newsletter*, 68.
- Chiang, M.-H. (2002). *Core competences for different civil service position ranks and establishment of training system*. Research commissioned by the Civil Service Protection & Commission.
- Cofsky, K. M. (1993). Critical Keys to Competency-based Pay. *Compensation and Benefits Review*, 25(6), 46-52.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
- Gardner, H., Feldman, D. H., & Krechevsky, M. (1988). *Project spectrum: Early learning activities, Project Zero frameworks for early childhood education* (Vol.

- 2). Teachers College Press.
- Guilford, J. P. (1988). Some changes in the structure-of-intellect model. *Educational and Psychological Measurement*, 48(1), 1-4.
- Jiang, T.-J. (2022). Is the civil service examination no longer popular? Reason behind the decline in the number of applicants for civil service examinations. *National Human Resources Forum*, 18.
https://www.exam.gov.tw/NHRF/News_EpaperContent.aspx?n=3778&s=45441&type=DD102593FDB1A032
- Kao, Y.-K. (2013). The analysis to the relationship between national examinations of civil servants and competence based test. *National Elite Quarterly*, 9(2), 1-22.
- Lee, D.-K. (2012). A study of examinations for foreign service and core competence. *National Elite Quarterly*, 8(1), 87-115.
- Lee, T.-W., Lee, C.-H., Hu, J.-P., & Huang, C.-L. (2012). *Research on university multiple admission pathways*. Research commissioned by the Research, Development and Evaluation Commission, Executive Yuan.
https://ws.ndc.gov.tw/001/administrator/10/refile/5644/3267/0058949_1.pdf
- Lu, C.-Y., Lu, C.-M., & Ou, T.-H. (1994). *Multifactor Aptitude Tests*. Chinese Behavioral Science Corporation.
- McClelland, D. C. (1973). Testing for competence rather than for “intelligence.” *American Psychologist*, 28(1), 1-14.
- Ministry of Examination. (2013). *The evaluation results of national examination occupational competency analysis promotion work*. Ministry of Examination.
- Ministry of Examination. (2020). *The physical checkup standard form of the air traffic controller category in the special examination for civil aviation personnel*.
https://wwwc.moex.gov.tw/main/controls/wHandEditorExtend_File.ashx?Fun=Pproperty&menu_id=335&item_id=4856&file_id=9808
- Office of Personnel Management. (2019). *Delegated examining operating handbook: A guide for federal agency examining offices*. https://www.opm.gov/policy-data-oversight/hiring-information/competitive-hiring/deo_handbook.pdf
- Peng, C.-P., Cheng, S.-C., & Lee, C.-T. (2017). *Feasibility of psychometric testing in civil service examination*. Ministry of Examination.
- Shih, N.-C. (2010). Lessons from competency theory for critical issues of national examination system design. *National Elite Quarterly*, 6(3), 17-35.
- Shih, N.-C., Tsao, R.-T., & Tsai, H.-C. (2009). The selection system of newly-hired civil servants at the central government of the USA, UK and Japan. *National Elite Quarterly*, 5(1), 13-34.
- Spearman, C. E. (1927). *The abilities of man, their nature and measurement*. Macmillan.

- Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. John Wiley & Sons.
- Sung, Y.-T. (2015). *Computerized adaptive career aptitude test for junior high school students (5th ed.)*. Research Center for Psychological and Educational Testing.
- Terman, L. M. (1916). The measurement of intelligence. In W. Dennis (Ed.), *Readings in the history of psychology* (pp. 485-496). Appleton-Century-Crofts.
- Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38(5), 406-427.
- Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago Press.
- Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, 2, 94.
- Tuan, M.-Y. (2006). *A comparative study of psychological tests used in personnel selection in Taiwan* [Unpublished master's thesis]. National Sun Yat-sen University.
- Vocational Training Bureau, Ministry of the Interior. (1985). *General Aptitude Test Guidelines*. Bureau of Employment and Vocational Training, Ministry of the Interior.
- Wu, M.-H., Chen, R.-H., & Chen, H.-Y. (2011). *New edition multiple talent aptitude test guidelines*. Chinese Behavioral Science Corporation.
- Wu, T.-R. (2012). The timing for adopting psychological tests in police officers examinations. *National Elite Quarterly*, 8(1), 117-151.
- Yeh, Y.-L., & Ting, Y.-K. (2021). Review and reflection on Taiwan's multiple admission pathway system. *Taiwan Educational Review*, 10(4), 1-9.
- Yu, M.-N. (2015). Issues about the item bank and test construction after the modified Examination Affairs Act. *T&D Fashion*, 211, 1-19.
- Yu, M.-N., Hsieh, J.-C., Chang, Y.-H., Li, P.-H., & Peng, Y.-C. (2016). *Public sector using psychometric testing and related legal backgrounds*. Civil Service Development Institute, Directorate-General of Personnel Administration.