Critical Considerations on Issues in Character Assessment[*]

Lee, GiYeon[**]

Seoul National University

ABSTRACT

Evaluation is essential for developing effective education and for introducing it to students. The same applies to character education as well, but evaluating character is not a simple task. Character is a value-oriented concept of 'a desirable image of human being,' constantly changes throughout one's life, is often affected by situations, and has multiple aspects. Regarding these characteristics, I discuss that interpretive tradition is more suitable than positivist tradition in evaluating character. When positivistic measurement participates as a pivotal role, serious problems would occur. It could lead character education into the mean of indoctrination, and students' characters could be distorted by concentrating only on some aspects. After all, it could make all of the efforts on character education useless that had been done before. Here I suggest principles of character assessment to retain its original purpose for improving character education and cultivating students' characters.

I. Introduction

The Character Education Promotion Act, which was began to be discussed on the occasion of a severe school violence in Daegu in 2011, passed the plenary session of the National Assembly with the consent of all the attendees on December 29th, 2014. Unlike the active support of the National Assembly, there are strong arguments between opposing and proposing sides on the Act in the press. The purpose and aims of the Act were mostly approved, but the ways and viability are still in question. The particular reason why the Act is criticized is because the contents of the Act are devoted to evaluation rather than education itself (Kim, 2015; Kim, 2016). The announcement by the Department of Education to reflect the results of character assessment into university entrance examination caused enormous controversy in the Act. While the Department's plan was abolished by the backlash, the influence is still affecting the private character education market (Huh, 2015). The cultivation of character, that should be the ultimate aim of education, is about to be 'commercialized.'

Though the Act is controversial, it doesn't mean that the evaluation itself is unnecessary. First, evaluation is necessary in linking the gap between intention and effect of education method. Education is generally defined as "the planned change of human behavior." (Hwang, 1998: 31) Although human behavior can be changed by almost anything that the person experiences, unplanned, random, or unintentional education precludes many elements in education. The planned education has a clear goal about what to develop through the education and it has a clear theoretical and empirical basis for how it can be developed. Evaluation can make the planned education possible by being a bridge between the intention and effect of education.

Second, the evaluation process can contribute to the development and the dissemination of great educational programs by verifying the effectiveness of the education. Without a systematic and scientific evaluation system, character education could be sporadic and faddish (Cunningham, 2005), and unsubstantiated programs could be introduced to students (as cited in Harrison & Davison, 2014: 2). For example, the Drug Abuse Resistance Education (D.A.R.E.), which was developed in the United States, in 1984, has been held in suspicion with its effect since 1992. Nevertheless, after three years, United Kingdom accepted the program and initiated it by police in schools (Berkowitz, 2014). The program is now classified as potentially harmful (Lilienfeld, 2007). In the end, students have been exposed to the poor program for quite a long time. If we avoid evaluation, there is no way to figure out how many conventionally being operated programs actually harm students' character, and the ideal of character education

for a whole-rounded person would remain nothing more than just a slogan (Hwang, 1998).[1]

Finally, evaluation itself can be an important part of character education. The ultimate goal of evaluation is to guide education to improve, and the ultimate goal of character education is to cultivate one's character. For this, it is important to make students reflect on their characters, and this valuable process of "Know thyself" could be done more systematically through evaluation (Galton, 1949: 179). The result of a long-term and continuous evaluation could be returned to students as a feedback, not as a report card, and students can review it to make improvements. Not only students but also teachers can be guided to enhance their teaching through the results of evaluation about school atmosphere, class culture, instruction skills, educational method, etc.

In order for the character assessment to be performed in accordance with its original purpose when the problems and the need coexist, it is essential to clarify issues around character assessment. Since the Act is already implemented, the influence would remain strong for a while. What we have to do now is to mull over how it can be performed with the least side effects, particularly over how to properly implement the evaluation. Rash judgment without deliberate considerations will detriment educational values. Therefore, in this paper, the need, current status, difficulties, possibility and more of the character assessment are examined to seek proper implementation of character assessment under the Character Education Promotion Act.

---

[1] For sure, the verifying process should be transparent and fair to convince school leaders or policy makers to adopt better educational strategies by simplifying the information without distortions (Berkowitz, 2014). In Korea, there is still a lack of awareness of the need to systematically verify effective and reliable character education programs. Besides, there are still some sorts of distrust in the current certification system. For example, Huh (2015) is concerned that as the Act would open up the possibility of involvement of outside experts, character education could be commercialized and become 'means of living' in the private education market. Involvement of experts can contributes to the quality of evaluation, but distrust in the certification process can make the merits be obscured. Objective and reliable scientific research, which opens the process of demonstration and opportunities to criticize and modify, can gain public's trust. Without this, unnecessary controversies would make students confused.

II. The Theoretical Ground for Character Assessment[2)]

A. The characteristics in the concept of character

In spite of high need of evaluation, the discussion for character assessment is inactive. It is because the concept of character is quite different from the academic knowledge, which used to be perceived as an object of evaluation. The concept of character (人性) has such an enormous depth and complexity, so it is not easy to make a single definition that everybody can agree on. However, 'evaluation' is basically "the process of judging something's quality, importance, or value." (evaluation, n.) For an objective evaluation, it is necessary to clarify whether the object has such a value or level in view of a certain standard, and a shared understanding of the evaluation object is essential for the clear standard. Therefore, in order to begin a discussion on character assessment, it is necessary to thoroughly examine what characteristics the concept of character has.

First, character is an evaluative and value-oriented concept. According to a survey operated by Korean Institute for Curriculum and Evaluation (KICE), students, parents, and teachers responded 'a moral person' first (63.4%), 'a well-rounded person' second (23.4%), 'creative person' third (6.6%), 'great citizen' fourth (0.3%) to a question about the goal of character education. As a result of combining the first and second responses, the ranking was as followed: 'a moral person' (88.8%), 'a well-rounded person' (53.9%), 'a great citizen' (28.8%), 'a creative person.' (26.7%) (Lee et al., 2011: 38-40) Through the survey result, we can apprehend that 'value-orientation' is at the core of students, parents, and teachers' understanding of character.

Historically, the same value-orientation can be found in what philosophers have searched for the meaning of character. When character literally referred as 'qualities of human being,' how to understand 'human' has been an essential question to every philosopher for a long time. Firstly, discussions in Eastern philosophy are as follows:[3)] In Confucianism, *Mencius* (孟子) thought that what makes human as a human is the

---

[2)] Character assessment is a part of the character education evaluation. The concept of 'evaluation' and 'assessment' are interchangeable. When examined in detail, however, specific purpose and distinctions can be varied. Here I chose 'evaluation' for character education as a general term, and 'assessment' for character to emphasize the importance of usage of various methods by various evaluators. Distinctions of each concept will be discussed in detail in this chapter.

[3)] I have to note that English translations used in this part are not enough to fully understand notions of Eastern philosophy.

morality of *Ren* and *Yi* (仁義), and suggested 'four beginnings' of the Good with "the feeling of commiseration," [惻隱之心] "the feeling of shame and dislike," [羞惡之心] "the feeling of modesty," [辭讓之心] and "the sense of right and wrong." [是非之心] (Kang et al., 2008: 16-17; Feng & Bodde, 1948: 169-70) *Chu His* (朱熹) generalized human′s original nature [本然之性] as a *Li* (理, principle) of the whole universe to find ontological grounds of human morality. He insisted human being intrinsically desires to be a moral person and has a possibility to achieve the ideal through the investigation of *Li* by approaching things [格物窮理] (陳來, 1992/1997). Meanwhile, in Buddhism, everything is tentative since it is constantly born and dead by relations [因緣]. As human also has a nature of emptiness [空性], we can reach Nirvana (解脫) by realizing the fact that there is no fixed self to cling to [無我] (Han, 2001). Finally, Taoism's *Tao* (道) or Self-so [自然] is the principle of this world and the foundation of human being. It is something unnamable and the very natural state of things without artificiality (Kang et al., 2008). Since the imperfection of this world comes from the excessive artificiality, human being has to brush aside the artificiality, including moral norms, and resemble the movement of *Tao*. Secondly, discussions in Western philosophy are as follows: Plato insisted that an ideal state of human being is achieved when our logic controls our spirit and desires. Only logical thinking from the universal point of view can recognizes the world of the *idea* beyond time-space constraints and mental and physical obstacles of us (Han, 2001). Aristotle thought that when logic and emotion function moderately, human can express human nature most perfectly and can lead the happiest and the greatest life (Arrington, 1998/2003). Meanwhile, Immanuel Kant insisted that we can be truly free when our transcendent-self controls the experiential-self, as it can release every emotional motivation (Kang et al., 2008). For him, the Good Will is the true nature of human.

A review of Eastern and Western philosophers' view on character reveals that individual philosophers' various perspectives on character were structured by their diverse value judgment. No matter how they describe the meaning of character, each of their view implicates the value judgment about what human ought to desire, which is, 'humanity.'[4]

---

[4] In psychology, the value-orientation of character has been considered inappropriate for a long time, and many scholars preferred to focus on personality, which is a rather value-neutral notion (Fowers, 2014). This trend was influenced by Gordon Allport a lot, who was a well-known personality psychologist. For example, when John Watson (1919) defined character as an evaluative notion, Allport insisted that with Watson's opinion, strictly speaking, researches on character belong to social ethics, not psychology (as cited in Allport, 1921).

Second, character is 'potentiality,' and character of a person constantly changes. In Eastern and Western philosophy, philosophers postulated something that human should desire (*Li* (理), Buddha Nature [佛性], *Tao* (道), *idea*, logic, Good Will, etc.), and every human was supposed to have a duty to cultivate it throughout their life. As no one is born with fully blown virtues, we cannot assure one's character of a moment is the completed status. It is always 'onward', which means it is heading to somewhere.

Third, character can never be detached from the influence of context. Kohlberg, who experienced and witnessed human immorality during the 2nd World War as a Jewish, was extremely against ethical relativism because it makes the absolute moral principle powerless. He preferred decontextualized and abstract reasoning that allows moral judgment from the universal justice. However, David Carr (1991) pointed out that virtues must be "contextually-specified and situationally-ordered," and "their meaning and expression of virtues are deeply embedded in the practices, customs, and expectations of communities," but still, it doesn't necessarily represent ethical relativism (as cited in Lapsley & Narvaez, 2006: 8). Hartshorne and May's (1928-1930) research also demonstrated how situational factors influence individual's moral behavior by observing different behaviors of the participated students at a test site, depending on whether the supervisor was an adult, how important the test was, or how big was the danger of consequence after the cheating was caught (as cited in Lapsley, 1996/2000). As people's different reactions to the same situation can be drawn in distribution graph, each has their own wide distribution of behaviors, and this stays consistent (Fleeson et al., 2014). It is almost impossible to completely separate the context from the discussion of character, and we should expand our consideration from individual's inner-self and primary relationship to groups and social relationship. Under the practical goal of character education, it is especially essential to consider the context of education and specificity of individual, comprehensively.

Fourth, character is a multifaceted concept, and contains various aspects. The Kantian ethics, where Kohlberg's theory is rooted in, regarded individual's character, trait, or virtue as obstacles to universal thinking so tried to exclude these. Similarly, Kohlberg's cognitive development tradition failed to realize the importance of character

---

However, even when discussing character in psychology, we should include evaluative and value-oriented level which can be represented by 'morality,' since it is a genuinely important part of human being and the discussion would be incomplete without it. Fleeson et al. (2014) insisted if we focus only on value-neutral personality, we cannot fully understand the core of our characters, suggesting differences between value-neutral personality and morality as follows: First, moral transgressions are rarer and more powerful. Second, social norms for behavior are more powerful for morality. Third, self-report tests about morality have less validity than self-report tests about personality in general. Fourth, moral behaviors usually implicate a conflict between self-interests and moral interests.

for a long time with alertness to ethical relativism (Walker, 2002). However, since Kohlberg's reason-oriented moral education was criticized that it was not enough to draw students' moral behaviors, discussion over other aspects perked up. For example, Martin Hoffman emphasized affection like sympathy, seeking 'Hot Cognition,' and Carol Gilligan emphasized the importance of care and relationship, insisting that there was no place for 'a different voice' in Kohlberg's theory. James Rest suggested Four Component Model (FCM) which does not distinctly separate areas of cognition, affection, and behavior. Anthony Blasi and William Damon emphasized the role of 'self' in developing moral behavior. Recently, Jonathan Haidt focused on intuition and unconsciousness, and Darcia Narvaez interpreted his research educationally and emphasized a series of ethical skills that can evolve 'moral expertise.' All of these aspects can be discussed as an important part of character, but focusing on only a few aspects would lead to a distorted view of one's character. The completed discussion on character would be possible when we synthetically apprehend how each aspect is performed and how they relate to each other in an individual as a whole.

B. The concept and approaches of evaluation

How can we evaluation character without omitting any of its characteristics? We generally think of a test or an exam when it comes to evaluation but regarding character's characteristics, evaluating character doesn't seem to accommodate with a test or an exam. Normally, we have some improper beliefs in evaluation: the most important thing in evaluation is to be better than others, numbers are always accurate and can make everything clear, only the multiple-choice test is reliable, or the accountability of test results is entirely student responsibility, etc. (Kang et al., 2012: 14-19) However, Kang et al. (ibid.) insisted that these negative view is due to 'mystical beliefs' and we should resolve those misunderstanding so evaluation can function for its original purpose. In fact, evaluation is not only about a 'test.' It has various meanings due to the perspective of an instructor.

When we classify what is referred to as 'evaluation' according to its purpose and characteristics, it can be divided into 'measurement,' 'evaluation,' 'assessment,' etc. (Hwang, 1998: 44-55) First, measurement is a process of numbering to specify the properties of objects (Seong, 2010: 93). It could be the most similar concept to evaluation of our perception. Since unquantified information could develop misunderstandings due to the ambiguity and vagueness, measuring could be highly efficient in enabling smooth and accurate communication. This method is rooted in positivist perspective. Positivism considers obvious, realistic, and observable things as objects of research (Pring, 2000/2015: 178-179). It attempts to discover general rules in social science to predict

human behaviors (as cited in Sanderse, 2015). According to this perspective, we can exactly clarify the differences of individuals since human behavior is also unchangeable and fixed.

However, the presumption about unchangeable and fixed being could be in question. Since character keeps changing as discussed above, conducting an evaluation with the positivist premise would lead to a narrow conclusion. Unlike measuring, 'evaluation' starts from the supposition that everything 'changes.' Also, it includes everything that related to education (curriculum, teacher, educational methods, environment, etc.) as objects and tries to evaluate these comprehensively (Hwang, 1998). In addition, 'assessment' is not about using a single method for a single object, but about collecting every possible data including both quantitative and qualitative one (ibid.). Evaluation and assessment are based on interpretive tradition, which is opposed to positivist tradition. Its main theory is that social science should be researched with its own method which is different from method of natural science. Human always gives 'meaning' based on their own 'interpretation' when interacting with something or someone. Therefore, to fully understand human behavior, we have to know not only the person's observable behavior but also his/her interpretation and intention (Pring, 2000/2015). Since we always interpret, it is almost impossible for humans to have an unprejudiced attitude toward an object (Gadamer; as cited in Cho, 2015a: 109).

Measuring presumes unchangeable substance, produces the result in numbers and denies the intervention of values. Considering the concept of measuring and the epistemological tradition that it is based on it, measuring should be used only as an substitution method in order to reflect the characteristics of character on. When measuring becomes the principal of evaluation, serious problems occur: character education would be devoted to indoctrination by teaching predetermined, monolithic concept of ideal of values and virtues; education and evaluation would concentrate only on cognitive domain, which is relatively easy to be quantified; evaluation would be considered as "quibbling, digging, scolding, judging, blaming, and classifying." (Hwang, 1998: 39) Eventually, it would be easy to reach a conclusion such as 'This is how many points your character gets.' (ibid.) The ultimate goal of educational evaluation, however, is to help education in any way, not to be the end state.

III. Practical Use of Character Assessment

A. Structure of character assessment

Then how character assessment can actually be operated? For actual improvement of character education, evaluation should include everything that could be an answer of "What could be changed for a better character education?" as objects of evaluation, and the range should cover from the start to the end of character education. Berkowitz (2014), who have thoroughly verified the effectiveness of character education programs, suggested 'Implementation Fidelity,' 'School Climate,' and 'Student Outcomes' as objects of evaluation. Here I re-categorized his standard and organized evaluation factors under 'the design and implementation of character education program' and 'the effect of character education program.'

1. Evaluation methods for the design and implementation of program

As mentioned above, evaluation should be operated for everything related to the educational activity. Though the ultimate goal of character education is the cultivation of individual's character, if we have a narrow sight by only focusing on it, it would distort the educational activity and suggest educationally undesirable remedies. An example of this problem is Herrnstein and Murray's (1994) research on 'Bell Curve' and their interpretation. When the researchers found that the distribution of subjects' IQ and the distribution of subjects' income are quite similar, they concluded that the African Americans had averagely low income because their intelligence was genetically low (as cited in Fisher et al., 1996: 6). Among the controversy the result brought, Fisher et al. (ibid.) insisted that it was not easy for low-income people, who were alienated from educational opportunities in the first place, to receive high grades on the test because the intelligence tests were usually based on knowledge learned at school.

Likewise, when only focusing on individual's character consequentialistically in character assessment, it is highly possible to grasp the result incorrectly. Therefore, before evaluating, it is necessary to first understand what kind of educational activities were done before evaluation, and how these relate to the social and cultural factors at the time of the evaluation. To accomplish these, we can evaluate 'the fidelity[5] and the

---

[5] 'The fidelity of the implementation' is about whether the program was implemented as planned. Even if a doctor prescribed a great medicine, we could not know the effectiveness of the medicine if the patient didn't take it well. Likewise, however well-designed a character education program is, we cannot verify its

quality[6] of the character education program' to identify if the program was properly designed and implemented, then evaluate 'the environment of character education program' to find out if the atmosphere of school, class, and society could affect positively on character education.

Program evaluation standards of Charagter.org (former Character Education Partnership, CEP) and The Collaborative for Academic, Social, and Emotional Learning (CASEL) can be good examples. First, Character.org (www.character.org) is a nonprofit, nonpartisan, and nonsectarian institution founded to develop effective character education in schools. It accepted Thomas Lickona's (1996) 'Eleven Principles of Effective Character Education' and selects the National Schools of Character (NSOC) and Promising Practices every year based on their evaluation standard. In addition, CASEL (www.casel.org) is a leading institution of Social and Emotional Learning. Following the publication of *Safe and Sound report* (CASEL, 2003), it publishes a guidebook every year to provide systematic standard for evaluating effective Social and Emotional Learning program. An example of its standard for SEL programs for preschool and elementary school is as below (CASEL, 2013: 19-30):[7]

\<Table 1\> Rating table for SEL Programs for preschool and elementary school

1. Program Design and Implementation	
Grade range covered	Participated students' grade levels.
Grade-by-grade sequence	Whether the program provides improved education as students' grade changes, whether educational contents are not overlapped, or whether there is omitted grade.
Average number of sessions per year	※ 'Session' is a set of activities designed to take place in a single time period.

---

effectiveness if schools don't implement it faithfully because of some practical problems like the lack of budget.

[6] 'The quality of the implementation' is about whether the program was designed and implemented well in the educational sense. For example, evaluation of teachers' attitudes or educational view, educational methods or content, etc. could be included here.

[7] Evaluation for * is in 3-point Likert scale, and others are on a pass or fail.

Classroom approaches to teaching SEL*	*Explicit SEL skill instruction*	Explicit education about SEL skills
	Integration with academic curriculum areas	SEL education through core academic subject
	Teacher instructional practices	Teacher level education though making positive class climate, etc.
Opportunities to practice social and emotional skills		Whether there are opportunities to practice SEL skills within the program, or outside of the program (real world).
Contexts that promote and reinforce SEL.	*Classroom beyond the SEL program lessons*	Whether the program provide context to reinforce SEL in classroom
	School-wide	Whether the program provide context to reinforce SEL in school
	Family	Whether the program provide context to reinforce SEL in home
	Community	Whether the program provide context to reinforce SEL in community
2. The Evidence of Effectiveness		
Grade range covered		Participated students' grade levels
Characteristics of sample		the grade levels, the geographic locations (urban, suburban, rural) where the studies were conducted, student race/ethnicity, and the percentage of students receiving free or reduced lunch
Study design		Quasiexperimental or Randomized Clinical Trials
Evaluation outcomes	*Improved academic performance*	e.g., grades, test scores

	Improved positive social behavior	e.g., working well with others, positive peer relations, assertiveness, conflict resolution
	Reduced conduct problems	e.g., aggressive or disruptive behavior
	Reduced emotional distress	e.g., depressive symptoms, anxiety, or social withdrawal

In addition, evaluation of 'the environment of character education program' enables us to have a bigger vision. Possible topics are teachers' language habits and behaviors, shared rules, the quality of relationships of school members, etc. Though it is included in the standard of CASEL, it can also be evaluated separately with tools like the Comprehensive Measuring School Climate (CSCI) invented by National School Climate Center (www.schoolclimate.org) or Association Test invented by Kamizono and Morinaga (2012).

2. Evaluation methods for the effect of the program

Each individual is the subject of character education, who is desired to show the effect of the education. Since the essential purpose of character education is to cultivate individual's character, showing how individual character has changed would be the most obvious way to discover the effectiveness of the character education program. Still, since character is an extremely complex concept, making a harsh judgment on one's character would bring serious side effects and change the character education into an immoral one. Therefore, when evaluating one's character, we always have to be cautious in the result.

As character has various aspects, evaluating on individual character could be so diverse due to evaluators' agenda. For example, as CASEL aims to students' academic, social, and emotional development, it makes the academic performance, positive social behavior, conduct problems and emotional distress as topics of evaluation. Or, Jennifer C. Wright (2014: 5-10) suggested using stimuli in evaluating character, organizing the evaluation frame with 'Measuring sensitivity to the presence of *virtue-relevant stimuli*,' 'Measuring recognition and generation of *virtue-appropriate responses*,' and 'Measuring dispositionality.' In recognition of Wright's standard covers the overall process of character, here I modified it a little bit and suggest an evaluation standard consists of 'the

general understanding of one's own character (before stimuli),' 'cognitive/emotional responses to stimuli,' and 'behavioral responses to stimuli.'

First, we can evaluate 'the general understanding of one's own character.' When we are to examine one's character, the simplest way to get an answer is to directly question the subject (or people around the subject) about their established understanding of his/her character. It is highly efficient and can be operated simply, but has a con that it is usually operated compactly without the context. The most common evaluating method would be self-report and other-repot. Self-report and other-report are usually in the form of a survey. They directly give questions to people about their character or give indirect questions using behaviors derived from the aspect or similar properties (Robins et al., 2009). For example, when evaluating 'diligence,' evaluators can present a sentence as 'I/he/she is diligent.' or 'I/he/she have/has been up all night a lot because I/he/she couldn't finish homework.' In addition to the Likert-scaled rating system, presenting open-ended questions could also be another way. 'Present an anecdote related to my/his/her diligence.' can be a simple example. Possible tools for this area are Value in Action Inventory of Strength (VIA-IS; Peterson & Seligman, 2004; www.viacharacter.org), Positive Youth Development (PYD; Geldhof et al., 2004), Ethical Skills Assessment Tools (Narvaez et al., 2001a-2001d), Social Skills Rating System (SSRS; Gresham & Elliot, 1990), etc.

Second, we can evaluate 'cognitive/emotional responses to stimuli.' Evaluation tools introduced above make subjects compactly express their character by ruminating their past experiences. On the contrary, this part is about evaluating subject's present status by presenting a specific stimulus and analyzing the responses of subjects. There are various types of stimuli: a scenario that describes a specific situation and asks possible action in there, a dilemma that presents a situation of conflict between values and evaluates subject's judgment and adaptability in there, a word that can stimulate associations, a combination of words, pictures, etc. Possible evaluation objects are also various: sensitivity or attitude toward stimuli, a judgment in a certain situation, behavior skills that can enable moral action, etc. Responses of the subject can be quantified through Likert scale, or coded qualitatively through an interview, writing, or presentation. Possible tools for this area are Racial and Ethical Sensitivity Test (REST; Brabeck et al., 2000), Moral Judgment Interview (MJI; Colby et al., 1983), Defining Issues Test (DIT; Rest, 1986), etc.

Finally, we can evaluate 'behavioral responses to stimuli.' Evaluation of behavior is usually operated through observation. Since changes in individual's behavior can be one ultimate goal of character education, capturing one's behavior reflecting his/her character would be a critical task. Behavior can be observed in an experimental situation and real situation. First, in the experimental situation, researchers can construct a scene, controlling variables explicitly. If we conduct experiments repeatedly switching the

context, it would be possible to figure out mechanisms of character from the various perspectives, and this would help us to understand character as a wide spectrum, not as a single dot. However, we cannot guarantee that we would find the same behavior in real life that was found in the experimental situation as one can make up his/her moral behavior when they are watched by observers. Since variables in a real situation are never simple or obvious as in an experimental situation, it is important to find their true behavior also in real life. Possible tools for this area are the interactive virtual reality vignette exercises used by Paschal et al. (2005) or Electronically Activated Recorder (EAR) used by Bollich et al. (2016)

B. Methods for character assessment

These tools introduced above can be classified in their forms: self-report, other-report, dilemma test, interview, observation, etc. Classifying tools in forms makes it easy to check each method's pros and cons so that it can let us know which kind of method is appropriate or inappropriate for a particular subject or a situation.

1. Self-report

Self-report is the most commonly used form in personality test or character assessment (Robins et al., 2009). Though it looks quite simple, a lot of costs, techniques and efforts are needed to construct even one question because it has to draw respondents' answers with only one sentence question. For example, sentences with 'should,' 'appropriate,' or 'would' all can draw different answers even when they contain the same content (Roberts, 2015). It is also possible that each respondent has different criteria on the same question. Therefore, researchers must make clear what data to get, construct plausible items, and suggest them with specific instructions. There are a lot of objects that can be evaluated with self-report. Questions can be constructed on individual habit, patterns of behavior, faith, attitude, preference, etc. Since evaluating one's inner self can only be answered by the respondent him/herself, self-report becomes the only method possible.

Robins et al. (2009: 226-28) gave self-reports' pros as follows: As self-report presents questions to someone who can truly look into his/her inner side, it can attain richer and more worthy answer than any other methods. And the use of it would be so efficient and cheap regarding time and budget. However, self-report has a fatal problem that we cannot sure if its result truly tells the respondent's character. Brown and Shelmadine's (1928) research showed this problem, compressively. When they conducted

self-report tests to students' characters, most of the students already had knowledge about what was the right behavior but knowing didn't guarantee what to do in real place observation. Moreover, in the test about honesty, honest students tended to rate their honesty low, but dishonest students tended to be generous to themselves, hiding their dishonesty. Through a series of experiments, researchers concluded that we should not completely trust the result of a self-report test.[8]

2. Dilemma test

Dilemma test has been actively used in evaluating moral judgment since Colby and Kohlberg's Moral Judgment Interview (MJI). It draws respondents' judgment and adaptability, using 'stories' containing conflicting values. Because of a strong pro that it can show specific contexts around moral issues, it is now variously used for many aspects not only for moral judgment.

Evaluation through dilemma test enables deeper understanding about the subject by widely catching their adaptability in a specific context, and it can handle the veridicality issues better as it tests indirectly. However, dilemma test also has a weakness that the kind of suggested stimulus can highly affect the test result. Since people normally can have faster and more accurate judgment to a familiar issue, if suggested dilemma reflects a certain culture, it can misevaluate people who do not belong to the culture. Franklin (1974) points out that most of the current dilemma test does not reflect circumstances of minorities so that it can underestimate their characters. As our society keeps changing into the multicultural society, we have to select stimulus for evaluation sensitively and seriously, and construct evaluation that can fairly treat people with different gender, religious, race, ethnicity, political stance, etc.

---

[8] Fowers (2014: 316-317) calls this problem 'Veridicality Issues' and explains it with three sub-elements. The first is the problem of social desirability response bias. On self-report, respondents tend to choose something depending on socially desirability rather than their true selves. Especially, as evaluation of character includes value determination about what is right or wrong, respondents hide their true selves when realized that their status is not appropriate. The second is the problem of self-enhancing positive illusion. People tend to have unrealistically positive impressions on themselves. Since there is always a possibility that what we believe about ourselves is not actually the truth, we may choose wrong answer to the question about our character, even unconsciously. The third is the problem of the confirmation bias. In self-report, respondents have to ruminate their past lives and abstract it shortly. During the process, exaggeration and distortion can happen and respondents can selectively recall memories that are favorable for their already established self-awareness. To overcome these problems, Fowers (ibid.) suggested removing the influence of social desirability statistically by including measures of social desirability, constructing questionnaires with a neutral tone, or making socially undesirable items more attractive.

3. Other-report

Other-report test is that makes people around the subject evaluate him/her. It is usefully when the subject is not appropriate to make self-report (e.g., when the object is a toddler), when the reliability of the self-report is not sufficient, or when the accuracy of the evaluation should be enhanced by being open to many raters (Hofsted, as cited in Robins et al., 2009: 260). Still, it can be used anytime when the self-report can be used.

This method can overcome the veridicality issues of self-report to some degree by producing rather an objective result. However, as we cannot tell one's self-awareness is actually true for his/her character, it is also hard to say that opinions of people around the subject would always be the truth. People in a close relationship with someone can have rather positive evaluation about him/her, and people in a bad relationship can have rather a negative evaluation. Also, sometimes a low score can be produced when raters had high expectation on the subject but the expectation wasn't fulfilled. One can hire experts who can observe the issue objectively as a third person to overcome these limitations and to enhance the reliability and quality of evaluation. Experts, however, would not have enough time to observe the subject thoroughly and the process would cost a lot. It is never easy to find the balance between the distance to the subject and objectivity. Therefore, we need to be open to a lot of raters' voices and have the attitude not to beg the question.

4. Interview

The interview is about getting information through verbal interaction, and it has a long history. The types of interview are as follows: 'Structured interview' is the interview with standardized questions and limited choices of answers. 'Unstructured interview,' on the other hand, is the interview approaching interviewees by communicating with them in depth, without fixed questions or answers. The structured interview is relatively objective because of the well-organized questions and consistent inquiry. But as the interviewer does nothing more than reading questions and writing down the interviewee's answers, it is not so different from the self-report. On the other hand, the unstructured interview can draw rather rich information because of unconstrained questions and answers.

The biggest pro of interview is that it can provide immediate feedback to the interviewee's answers since they interact in person. Especially in the unstructured interview, the interviewer can draw the ultimate reason of answers from the interviewee, by constantly asking 'why.' This enables to understand the interviewee in depth. However, evaluation with interview has limitations that it costs much more time and

budget than self-report, and that interviewer's quality decides the success of the evaluation (Cho, 2015b).

5. Observation

Observation is to capture the behavior of the subject directly, and also has a long history like interview. The type of observation can be classified by the role of observers. One can be 'the complete observer' when completely being separated from subjects, 'the observer as participant' when showing his/her presence on the scene but only observing, 'the observer as participant' when actively interacting with subjects, or 'the complete participant' when completely being integrated into the community of subjects (ibid.: 127-128).

Strong pros of observation are that it can directly check one aspect of character (Robins et al., 2009), and it isn't affected by response bias when compared with other methods. On the other hand, there are also limitations. The most severe problem is that observer's subjective recording and interpretation could contaminate the result and decrease the reliability. Therefore, researchers have to examine thoroughly what to evaluate and what could be the criteria for it through enough theoretical and practical consideration. It is important to consent about the interpretation process beforehand (ibid.). Plus, it needs quite a lot of budget, and subjects can change their behavior as they are being watched by an observer.

6. Other methods

Recently, there suggested methods that can 'directly' look into human unconsciousness and process of thought without any proxy. For example, Greenwald et al. (1998) applied Carl Jung's (1990) Association Test to develop Implicit Association Test (IAT). It is for finding out respondents' unconscious prejudices, presuming that one's favored objects are connected to positive words and one's dislikable objects are connected to negative words.[9] With this tool, Greenwald et al. (ibid.) indirectly measured

---

[9] The experiment consists of 5 stages. In the first stage, participants were asked to press right button if the presented word is pleasant and left button if the word is unpleasant. In the second stage, participants were asked to press right button if the presented word referred to what he/she liked and left button if the word referred to what he/she disliked. In the third stage, words that were presented first and second stages showed up randomly, and participants were asked to press right button if the presented word is pleasant or referred to what he/she liked and left button if the word is unpleasant or referred to what he/she disliked. The fourth stage is the opposite of the first stage. Participants were asked to press right button if the presented word is 'unpleasant' and left button if the word is 'pleasant.' In the fifth stage, participants were asked to press right button if the presented word is 'unpleasant' or referred to what he/she liked and left button if the word is

respondents' attitudes toward flowers and insects. Using IAT, there are researches about discriminative attitude toward race and sex (McConnell & Leibold, 2001), and about self-esteem and self-concept (Greenwald & Farnham, 2000). Since exaggeration or distortion can happen when people explicitly express their opinions, IAT would be useful to look into one's unmannered mind.

Also, the attempt to see one's mind with Neuroscientific methods is being actively developed. Hyemin Han (2016) suggests using Neuroimaging method to capture one's inner mechanism of virtue by, for example, comparing changes in a moral exemplar's brain and a normal person's brain under a certain stimulus. He explains that evaluating through 'proxy' has clear limitations: in self-report, we cannot see the substructure as it is hidden behind a behavior, and biased responses are common due to respondent's subjectivity. In this sense, Neuroscientific methods can be so useful if used well. Nevertheless, as there are still tensions between Neuroscience and moral philosophy, and it would cost a lot to use Neuroimaging, there should be more theoretical and practical research to apply the technique popularly.

IV. The Danger of Character Assessment

As discussed so far, there are tons of ways to evaluate character. These ways have both pros and cons, and appropriate method of each aspect is different from others. To use the most appropriate method at the most appropriate time, we need a lot of time, budget, and effort. On this account, some people choose to take easier and more efficient 'shortcut': using only a self-report and believing the result definitely. However, since a hasty character assessment can put our students in serious danger, we must avoid the choice of taking a shortcut.

A. The danger of assessment on value-oriented character

The most serious problem of character assessment is that the process and result of evaluation are easy to be misused, as it is hard to have consented criteria for character. Morality is based on the agreements of people, not in physical entities (Haan, 1982). Though the faith for universal moral principle across time and place has been proved

---

'pleasant' referred to what he/she disliked. If participants' response time in the third stage and the fifth stage differs a lot, it would mean that pleasant words are connected to likable objects and unpleasant words are connected to dislikable objects in participants' minds (Greenwald et al. 1998: 1464-1467; Lee, 2007: 9-10).

through a long history, norms in a specific scene of life could be diverse. In addition, the influence of context in cultivating character is crucial. Therefore, if one group determines all of the rules, contents, and evaluation criteria of character education, certain intentions can be included during transferring values, making character education cramming or indoctrination. In Japan, where people experienced these side-effects during the 2nd World War, prohibited numerical evaluation on any value-related content since the war (Kamizono, 2008).

Among the eight problems of virtue evaluation suggested by Howard Curzer et al. (2014), three belong here. First, it is hard to agree in moral theory. Historically, a lot of moral theory existed, but it is still controversial to say one theory is superior to others. If criteria change due to pursuing moral theory, it would be almost impossible to have consented criteria for evaluation. Second, it is hard to agree in actual practice. It is not only because ethicists pursue different moral theories, but also because they draw different conclusions from the same theory. In addition, the human world is so complicated that right and wrong cannot be decided that easily. If there is no clear answer, how can we evaluate a subject's answer? Third, the differences in knowledge on the stimulus used in the evaluation can lead to a distorted result. Everyone has their own spectrum of experiences, and it is natural to respond elementarily to questions about issues that he/she is not used to. This elementary answer can be misinterpreted that he/she has a low level of character.

If someone can set criteria for character assessment, who is he/she and what gives him/her the authority? Roberts (2015) suggests a solution to this problem: We need to give authority to elites who are verified through experience, reflection, historical understanding, and a long period of discipline. Still, this cannot be a perfect solution. For instance, developers of Adolescent Intermediate Concept Measure (Ad-ICM) organized their rating team with experts, but there was a response bias with gender in the result. The possible cause was pointed out that women might have contributed more when making stories and questions for the research (Thoma et al., 2013). It suggests the difficulty in evaluating value-oriented concept objectively.

B. The danger of assessment on contextual character

As discussed above, the reason why character's value-orientation could cause a problem is because the context has a strong influence in cultivating character. It is applied not only to macroscopic culture, but also to every small scene of our lives. In Hartshorne and May's research, people never acted consistently across different situations. As discussed by many scholars, it doesn't mean that individual's character doesn't exist, but that individual character exists as a wide spectrum, not as a dot.

Therefore, the attempt to exclude the context and show evaluation result fragmentarily can distort one's character by omitting social and cultural factors. There is an actual example that after a student honestly confessed his/her impulse, the answer was coded fragmentarily, and the result concluded that the student had a mental disorder (Park, 2015). With this kind of evaluation, one's long internal process for the response would become nothing.

Among the eight problems of virtue evaluation suggested by Curzer et al. (2014), one belongs here. People are not consistent across different situations. One can consume a couple of moral theories at the same time even when it is paradoxical, and can have different conclusion from the same theory due to situations. The complexity of modern society's moral issues also contributes to this phenomenon.

Likewise, character should always be considered with context. We cannot truly evaluate one's character by simply comparing the input and output of character education and quantifying the level of concord (Kvernbekk, 2014). In addition, evaluation that focus only on the universal principle is possibly be nothing more than simply checking knowledge. Rather, using everyday moral with context would encourage evaluation with various methods.

C. The danger of assessment on changeable character

One's character changes not only across situations but across time. As discussed in philosophers' views on the concept of character, character describe what human being should ultimately pursue, and this means that ourselves of now are not someone already achieved it, but someone keeps heading toward the ideal status. One's character is not completed and fixed in a certain moment. Therefore, we can never conclude that a status captured in one moment is the final status of one's character. If we use evaluation result as if it is everything of the student's character, it is denying his/her possibility to develop and will be nothing but a labeling effect.

D. The danger of assessment on multifaceted character

Even if character is not something changes across time and space, it is still impossible to evaluate it fragmentarily because character is such a complicated concept with various aspects. Whether it is a reason, affection, sociality, skills, or virtues, all of these are only a part of one's character. Moreover, character is not a simple sum of every element, but each element keeps producing new meanings by interacting with each other. Therefore, an attempt to measure this complicated and comprehensive concept in a

fragmentary way is not only inappropriate but also failing to recognize plurality and incommensurability of moral considerations (Haldane, 2014).

Among the eight problems of virtue evaluation suggested by Curzer et al. (2014), three belong to here. First, knowing moral theory is not enough. Most of the moral theories can have both grounds of pros and cons to the same moral question, and some people use moral theories to draw a wrong choice. Therefore, deciding which moral theories to pursue is only a small part of the whole moral development, and ability of decision-making, etc. should be considered together. Second, various aspects of character of individual develop unevenly. One can have fully developed virtue of care but still have low level of wisdom, and the other can have fully developed virtue of bravery but still have a low level of honesty. It is not only about virtues. One can have a high moral sensitivity but not enough power of execution by the lack of skills, and the other can have great motivation but make wrong choices by poor judgment. Since it is almost impossible to develop evaluation methods that can cover all of the aspects of character, it is possible to misunderstand someone's character with a method focused on only a few aspects. Third, people's explicit pursues are not always the same as their implicit pursues. As pointed out by Jonathan Haidt, people sometimes make decision by their intuition and explicitly justify it with their reason. Consciousness and unconsciousness should be considered at the same time because both affect the function of individual character.

If we conduct evaluation focusing only one part of character, it will arouse a question whether the result of evaluation is actually the subject's true character. When someone got 4 points of care and 5 points of justice, does he/she have lower level of character than someone who got 6 points of care and 7 points of justice? Or can we say that the latter 'actually' care others better than the former? Character assessment result about one aspect that is produced in simple number cannot fully describe one's character as a whole.

V. Principles for Character Assessment

Most of the problems around character assessment happen when the evaluation conducted only to produce results and to use the results conclusively. Are we trying to classify 'bad' character only to point the finger at the subject? If the evaluation is used to categorize students who have a 'bad' character, how can it help those students and what would be their future life? The formative evaluation also deprives of students' autonomy. Kant said that our fundamental moral duty to our students is to treat them respectfully as a 'person of equal worth.' Therefore, for a student of equal worth, teachers should help them to plan their own character development, reflect on their past behavior, and

critically evaluate it. Result-deriving evaluation, however, lacks the respect for a student of equal worth by subordinating them to predetermined evaluation criteria, and increasing their passivity (ibid.: 10-12). With this point of view, Jo (2013) insisted that developing a model of character assessment could be a necessary evil, because when the model is established, student would fix themselves into the 'customized character.'

These problems are deadly serious as evaluation, which should be not only an end point but also a re-starting point, can make the educational effects from the beginning disappear. As the ultimate purpose of character assessment is the development of character education, the result of evaluation should be the scaffold for the new start, not a completion. Therefore, to prevent these problems, we should protect morally righteous character education by generating principles for evaluation.

A. We always should be open to questions and critiques.

We should have open minds in determining and modifying the design, process, and result of character assessment. The beginning of education is to teach what we believe to be right, but the end is to leave it to students' judgments. That can be a true character education, which respects students as autonomous beings. We should always be open to critics and modification with keeping in mind the fallibility, and do the best to let students pursue what they believe to be valuable. Although character assessment is easy to be something immoral due to its wrong usage, it could be acceptable if it is used in the way that does not interrupt students developing their autonomy and rationality. It is the only way that we can uphold Kant's maxim (Siegel, 1980). For this, not only the usage of various evaluation methods but also the activity of various evaluators is important. According to Berkowitz (2014), in the time of value education, evaluation on character education program used to have principals as raters, and this naturally brought disagreements with teachers. Since all people have different perspectives and have possibilities to be wrong, we should be open to as many opinions as possible to avoid the abuse of character assessment results.

In fact, character assessment is a natural, inevitable, unconscious, and automatic process. Human being is a reflective being, and our self-reflection is based on our self-judgment. All of these processes can be character assessment of our own. In addition, as a social animal, human being builds relationships with others and evaluates their characters during the process. All of these processes could also be character assessments of others. However, as human beings are formed upon diverse experiences, we would never be free from our own prejudices. Even if we could exclude all of those prejudices, the possibility of objective evaluation is still far-off because there is no way to be sure of what we

observe in one moment is the truth. Thus, to get a more reliable result from character assessment, we need some 'scientific' methods.

However, there could be a question whether a scientific research on character is possible. I already concluded that positivistic measurement is not appropriate for character, but unquantified result could mean the lack of objectivity. When we decided to minimize the use of the method of measuring, is objective research on character possible?

It depends on how we approach the notion of 'objective.' According to Haan (1982), morality is a value-oriented concept based on people's social agreements. Therefore, when 'scientific' means value-neutrality, researches on character cannot be scientific. On the contrary, when 'scientific' means "impartially submitting all formulations to the full reality of people's moral consensuses and interactions in everyday life," researches on character can be scientific (ibid.: 1104). This alternative perspective on the 'scientific objectivity' can also be found in discussions of Standpoint Epistemologists. According to Harding (1992), so-called 'objectivity' is nothing more than a superstitious fiction, and it makes people pursue their unexamined faith without serious considerations. From their perspective, all knowledge is socially situated. It is impossible to delete the trace of the creator from the knowledge because there is no knowledge that is perfectly detached from social context. If we start from this point of view, what we can do for an objective research is to see the issue from as various perspectives as possible. It enables rather 'strong objectivity' through 'strong reflexivity.'

Character assessment also cannot be objective in the sense of deleting all contexts and values around the evaluation process, but can be objective in a sense of accepting as many perspectives as possible and never stopping to find the better alternative. What we need is to understand that it is impossible to evaluate one's character perfectly, and to have a modest attitude that what we do is only providing some help for a cultivation of one's character.

B. Character assessment should be the part of the educational process.

The term character education 'program' is often used followed by the boom of character education. The term 'program' has an impression of short-term and one-time, but character education can never be such a thing. Character education should be continuous and overall as one's character is cultivated throughout his/her life, thus evaluation also should be something constant. It should not be detached from the educational process, and its result should continuously be delivered to students as a feedback. Evaluation that we usually think of, like a mid-term or final-term exam, is

operated as a large event with a long interval and often focuses on its result to use it for class placement, seat placement in class, or university entrance exam. However, evaluation only for evaluating has no single meaning as education. Developing a great evaluation method is important, but to use it for a great purpose is as important. We should never forget that the ultimate goal of character assessment is for better character education.

C. Character should be evaluated with various methods.

The most important understanding for character assessment is that there is no perfect method like a master key. No matter how high reliability is, how great the scholars who developed it were, only one research or a method is not enough to evaluate even a single virtue (Fower, 2014). Each aspect of character has an appropriate method for its own content and characteristics. For example, questionnaires made with scenarios are appropriate to evaluate fairness, and interviewing about a complex issue is appropriate to evaluate open-mindedness (Peterson & Park, 2004: 439-440). Therefore, when we evaluate character, we should cooperate with scholars from various fields, and accommodate a lot of various methods and perspectives.

D. Assessment should cover diverse aspects of character and character education.

As the purpose of character assessment is the development of character education, the evaluation process should cover everything that answers a question "What could be changed for a better education?," and the range should cover from the start to the end of character education. Evaluation that focuses only on the individual's character can attribute responsibility of the result completely to students, which would negatively affect his/her character eventually. Therefore, we should pursue more rough and comprehensive evaluation by constructing the range of evaluation wider (Ellenwood, 2014). To do so, the evaluation of the design and implementation of character education and the evaluation of the effect of character education should be synthetically performed. The same principle should apply for each. Evaluation of the design and implementation of character education should include everything that can affect one's character like the educational interventions, the environment, the evaluation process itself, etc. Evaluation of the effect of character education also should not narrowly focus on a single aspect, but be conducted for character in general.

E. The outcomes of assessment should be qualitative and comprehensive.

Though the quantitative measurement of character is often used for its convenience and efficiency, this type of evaluation cannot describe one's character without distortion as it narrows the view of evaluator and omits a lot of stories (ibid.). Of course, it is true that a narrow evaluation is more useful in demonstrating the effectiveness of educational intervention. However, it limits the scope of character education a lot and makes it abstract, focusing only on what students don't know (ibid.). A qualitative evaluation method like portfolio would enhance the rich and deep learning of complex issues. Metaphorically speaking, character is a third or fourth dimensioned complex lump, not a single dot. If we try to produce result in numbers or a fragmentary sentence with greed to show the result explicitly and visibly, it would cover only a tiny part of the lump. As the American author William Bruce Cameron said, "Not everything that can be counted counts. Not everything that counts can be counted."

VI. Conclusion

So far I examined the need of character assessment, the concept and characteristics of character, proper evaluation approach, methods that can be used for each aspect, possible problems of character assessment, and principles to prevent those problems.

Though a significant amount of time has passed since the Act was activated, character assessment has not been established in schools because of a lot of practical problems. In fact, that 'evaluation should not be conducted only for the result, and the result should always be used for improvement of education,' which is an essence of this paper, is something that has always been emphasized in the foreword of introductory of educational evaluation. Although this is the most basic principle of every evaluation, it has not been followed well due to practical difficulties. Above all, one of those difficulties is that a high expertise is needed to conduct character assessment. All teachers are experts who are prepared with values and competencies for character education, but verifying effectiveness is a different task that requires different knowledge and techniques. In addition, overburdened tasks of teachers are chronic problems in schools. If we include character assessment, teachers will experience serious distress. Particularly, character assessment needs full attention for each student. With current educational system in Korea where one teacher handles large amount of students, it would bring nothing but a affliction.

However, since each school would set specific educational goal for its own, it would be better if someone who is a member of a school to establish an evaluation plan that is appropriate for the school. Therefore, researchers in the field of moral and ethics education can attribute to help teachers conducting character assessment more easily by constructing a 'standard module.' If researchers examine a series of well-designed character education programs and evaluation methods and make a list of them, educators could use it effectively when planning and performing their own character assessment. Overseas, scholars like Berkowitz and Bier (2005), institution like CEP or CASEL have already preceded this task, but still, there is no such data available yet in Korea.

To make a character assessment a strong supporter, not an obstacle of character education, theoretical and practical research should cooperate accordingly. This would solve the negative perception of character assessment and enhance the performance of it. I hope this paper would contribute to a part in the great task.

References

Allport, G. W. (1921). Personality and character. *Psychological Bulletin, 18*(9), 441.

Alexander, H. A. (2016). Assessing virtue: measurement in moral education at home and abroad. *Ethics and Education, 11*(3), 310-325.

Arrington, R. L. (2003). *Western Ethics* (Translated into Korean by S. H. Kim). Seoul: Seogwangsa. (Original work published 1998)

Berkowitz, M. (2014). Aligning Assessments in Character Education: Integrating Outcomes, Implementation Strategies, and Assessment. Paper presented at the Conference of Jubilee Centre, Oriel College, Oxford, January 9–11, 2014.

Berkowitz, M. W. & Bier, M. C. (2007). What works in character education. *Journal of Research in Character Education, 5*(1), 29.

Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook 1. Cognitive Domain.* New York: David Mckay.

Bollich, K. L., Doris, J. M., Vazire, S., Raison, C. L., Jackson, J. J., & Mehl, M. R. (2016). Eavesdropping on character: Assessing everyday moral behaviors. *Journal of research in personality, 61*, 15-21.

Brabeck, M. M., Rogers, L. A., Sirin, S., Henderson, J., Benvenuto, M., Weaver, M., & Ting, K. (2000). Increasing ethical sensitivity to racial and gender intolerance in schools: Development of the Racial Ethical Sensitivity Test. *Ethics 2000 Behavior, 10*(2), 119-137.

Brown, F. J., & Shelmadine, M. (1928). A critical study in the objective measurement of character. *The Journal of Educational Research, 18*(4), 290-296.

Carr, D. (1991). *Educating the virtues: An essay on the philosophical psychology of moral development and education.* London: Routledge.

陳來. (1997). *宋明理學* (Translated into Korean by J. H. Ahn). Seoul: Yeomoonseowon. (Original work published 1992)

Cho, Y. D. (2015a). *Qualitative research methodology: theory.* Seoul: Geun-sa/DreamPig.

Cho, Y. D. (2015b). *Qualitative research methodology: practice.* Seoul: Geun-sa/DreamPig.

Colby, A., Kohlberg, L., Gibbs, J., Lieberman, M., Fischer, K., & Saltzstein, H. D. (1983). A longitudinal study of moral judgment. *Monographs of the society for research in child development,* 1-124.

Cunningham, C. A. (2005). A certain and reasoned art: The rise and fall of character education in America. In D. K. Lapsley, & F. C. Power (Eds.), *Character psychology and character education* (pp.166-201). Notre Dame, IN: University of Notre Dame Press.

Curzer, H. J., Sattler, S., DuPree, D. G., & Smith-Genthôs, K. R. (2014). Do ethics classes teach ethics?. *Theory and Research in Education, 12*(3), 366-382.

Ellenwood, S. (2014). Measuring Virtue Better, A Little Later and Little Rougher. Paper presented at the Conference of Jubilee Centre, Oriel College, Oxford, January 9–11, 2014.

Evaluation. (n.) In *Cambridge English Dictionary*, Retrieved from http://dictionary.cambridge.org/dictionary/english/evaluation#translations. Accessed 12 April. 2017.

Feng, Y. L. & Bodde, D. (1948). *A Short History of Chinese Philosophy.* New York: Macmillan.

Fischer, C. S., Hout, M., Jankowski, M. S., Lucas, S. R., Swidler, A., & Voss, K. (1996). *Inequality by design.* na.

Fleeson, W., Furr, R. M., Jayawickreme, E., Meindl, P., & Helzer, E. G. (2014). Character: The Prospects for a Personality-Based Perspective on Morality. *Social and Personality Psychology Compass, 8*(4), 178-191.

Fowers, B. J. (2014). Toward programmatic research on virtue assessment: Challenges and prospects. *School Field, 12*(3), 309-328.

Franklin, A. J. (1974). The Testing Dilemma for Minorities.

Gadamer, H. (1989). *Truth and method (2nd ed.).* New York: Crossroad.

Galton, F. (1949). The Measurement of Character. In Dennis, Wayne (Ed), *Readings in general psychology,* New York, NY, US: Prentice-Hall, Inc., 435-444.

Geldhof, G. J., Bowers, E. P., Boyd, M. J., Mueller, M., Napolitano, C. M., Schmid, K. L., ... & Lerner, R. M. (2014). The creation and validation of short and very short measures of PYD. *Journal of Research on Adolescence, 24*(1), 163-176.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of personality and social psychology, 79*(6), 1022.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology, 74*(6), 1464.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual. American Guidance Service.*

Haan, N. (1982). Can research on morality be "scientific"?. *American psychologist, 37*(10), 1096.

Haldane, J. (2014). Can Virtue Be Measure?. Paper presented at the Conference of Jubilee Centre, Oriel College, Oxford, January 9–11, 2014.

Han, H. (2016). How can neuroscience contribute to moral philosophy, psychology and education based on Aristotelian virtue ethics?. *International Journal of Ethics Education*, 1-17.

Han, J. K. (2001). *The Eastern and Western Understanding of Human Being.* Seoul: Seogwangsa.

Harding, S. (1992). Rethinking standpoint epistemology: What is "strong objectivity?". *The Centennial Review, 36*(3), 437-470.

Harrison, T., & Davison, I. (2014). Assessing Interventions Designed to Improve Understanding of Virtues. Paper presented at the Conference of Jubilee Centre, Oriel College, Oxford, January 9–11, 2014.

Hartshorne, H., & May, M. (1928). *Studies in the nature of character Vol. Ⅰ, Studies in deceit.* New York: Macmillan.

Hartshorne, H., May, M., & Maller, J. B. (1929). *Studies in the nature of character Vol. Ⅱ, Studies in self-control.* New York: Macmillan.

Hartshorne, H., May, M., & Shuttleworth, F. A. (1930). *Studies in the nature of character Vol. Ⅲ, Studies in the organization of character.* New York: Macmillan.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment. Association for Supervision and Curriculum Development*, 1250 N. Pitt Street, Alexandria, VA 22314.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: The reshaping of American life by differences in intelligence.* New York: Free.

Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Developmental Psychology, 8*, 149-162.

Huh, J. R. (2015). *An Analysis on Newspaper Evaluations of Character Education Promotion Act, Journal of Human Rights & Law-related Education Researches, 8*(3), 175-199.

Hwang, J. G. (1998). *Hak-Gyo-Hak-Seup-Gwa Gyo-Yuk-Pyeong-Ga.* Paju: KYOYOOKBOOK.

Jo, J. W. (2013). *A Study on the Improvement of Admission Officer System in the University Entrance - Personality Evaluation in Center-.* Marster's dissertation, University of YoungNam.

Jung, C. G. (1910). The association method. *The American journal of psychology, 21*(2), 219-269.

Kamizono, K. (2008). Reticence towards Moral Lessons in Japanese Schools (道徳授業における控えめな立場: 岐路に立つ日本の道徳授業). *長崎大学教育学部紀要: 教育科学, 72*, 1-12.

Kamizono, K., & Morinaga, K. (2012). Assessment by Association Method of a Moral Education Lesson on a Local Topic in a Mixed-age Class. *長崎大学教育学部紀要: 教育科学, 76*, 1-16.

Kang, S. B. et al. (2008). *Character Education*. Paju: Yangseowon.

Kang, S. H. et al. (2012). *Modern Educational Evaluation*, Paju: KYOYOOKBOOK.

Kim, K. S. (2015). Culture: Problems and Suggestions of "Character Education Improvement Law." *Korean Thought and Culture, 78*, 459-486.

Kim, N. (2016). A Critical Consideration on the Idea and Implication of the 'Character Education Promotion Law.' *Studies on Life and Culture, 39*, 55-102.

Kvernbekk, T. (2014). On the Possibility of Interventions Aimed at Improving Character. Paper presented at the Conference of Jubilee Centre, Oriel College, Oxford, January 9–11, 2014.

Lapsley, D. (2000). *Moral psychology*. (Translated into Korean by Y. L. Moon). Seoul: Jung-Ang-Jeok-Seong. (Original work published 1996)

Lapsley, D. K., & Narvaez, D. (2006). Character education. *Handbook of child psychology*.

Lee, J. H. (2007). Sex Differences of Gender Category Representation in Implicit Association Test. *The Korean Journal of Woman Psychology*. 1-20.

Lee, M. J. et al. (2011). *A study on how to activate character education through subjects education and creative extracurricular activities*. Seoul: KICE.

Lickona, T. (1996). Eleven principles of effective character education. *Journal of moral Education, 25*(1), 93-100.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*(1), 53-70.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of experimental Social psychology, 37*(5), 435-442.

Narvaez, D. with Endicott, L., Bock, T., & Mitchell, C. (2001a). *Nurturing character in the middle school classroom: Ethical Action*. St. Paul: Minnesota Department of Children, Families and Learning.

_____ (2001b). *Nurturing character in the middle school classroom: Ethical Judgment*. St. Paul: Minnesota Department of Children, Families and Learning.

_____ (2001c). *Nurturing character in the middle school classroom: Ethical Motivation*. St. Paul: Minnesota Department of Children, Families and Learning.

_____ (2001d). *Nurturing character in the middle school classroom: Ethical Sensitivity*. St. Paul: Minnesota Department of Children, Families and Learning.

Park, S. H. (2015, November 27). An honest confession in a character assessment led to a label as problematic student. *JTBC*. Retrieved from http://news.jtbc.joins.com/article/article.aspx?news_id=NB11105703.

Paschall, M. J., Fishbein, D. H., Hubal, R. C., & Eldreth, D. (2005). Psychometric properties of virtual reality vignette performance measures: a novel approach for assessing adolescents' social competency skills. *Health education research, 20*(1), 61-70.

Peterson, C., & Park, N. (2004). Classification and measurement of character strengths: Implications for practice. In P. A. Linley & S. Joseph (Eds.), *Positive Psychology in Practice* (pp. 443-446). Hoboken, NJ: John Wiley & Sons, Inc.

Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification.* Oxford University Press.

Pring, R. (2015). *PHILOSOPHY OF EDUCATIONAL RESEARCH (2nd Ed.).* (Translated into Korean by Kwak, D. J. et al.). Seoul: Hakjisa. (Original work published 2000).

Rest, J. R. (1986). *Manual for the defining issues test.* Center for the Study of Ethical Development, University of Minnesota, Minneapolis.

Roberts, R. C. (2015). The normative and the empirical in the study of gratitude. *Res Philosophica, 92*(4), 883-914.

Robins, R. W., Fraley, R. C., & Krueger, R. F. (Eds.). (2009). *Handbook of research methods in personality psychology.* Guilford Press.

Sanderse, W. (2015). Aristotelian action research: its value for studying character education in schools. *Educational Action Research*, 1-14.

Seong, T. J. (2010). *Modern Educational Evaluation.* Seoul: Hakjisa.

Siegel, H. (1980). Critical thinking as an educational ideal. In *The Educational Forum, 45*(1), 7-23. Taylor & Francis Group.

The Collaborative for Academic, Social, and Emotional Learning. (2003). *Safe and Sound: An Educational Leader's Guide to Evidence-Based Social and Emotional Learning (SEL) Programs*

_____ (2013). 2013 *CASEL GUIDE: Effective Social and Emotional Learning Program (Preschool and Elemantary School Edition).*

Thoma, S., Derryberry, W. P., & Crowson, H. M. (2013). Describing and testing an intermediate concept measure of adolescent moral thinking. *European Journal of Developmental Psychology, 10*(2), 239-252.

Walker, L. (2002). Moral exemplarity. In W. Damon (Ed.), *Bringing in a new era in character education* (pp. 65-83). Stanford, CA: Hoover Institution Press.

Watson, J. (1919). *Psychology from the Standpoint of a Behaviorist*, Phila.: Lippincott.

Wright, J. C. (2014). Can Virtue Be Measured?. Paper presented at the Conference of Jubilee Centre, Oriel College, Oxford, January 9–11, 2014.

Author

Lee, GiYeon
Seoul National University
ibc6550@snu.ac.kr