

RESEARCH ARTICLE

Exploring the Effects of Overconfidence Bias and Hard-easy Effect in Self-monitoring of Biological Concept Test

Hyojeong Hwang · Minsu Ha · Eunju Park^{*}
Kangwon National University

생물개념평가의 자기모니터링에서 과신편향과 어려움-쉬움 효과 탐색

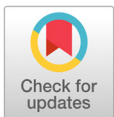
황효정 · 하민수 · 박은주^{*}
강원대학교

^{*}Corresponding Author: skyblueej@kangwon.ac.kr

ABSTRACT

This study explored the overconfidence bias and hard-easy effects of students' biological concept levels and self-monitoring accuracy. To this end, 361 high school first and second grade students participated in the study. As a consequence of this study, there were differences between genders or years in the four types of knowledge-confidence and significant differences between genders in specificity. As a result of clustering model analysis, hard-easy effects were found in a large number of students, and there were many cases in which the tendency to overconfidence was noticeable. Based on these findings, it was suggested that students should be educated to recognize and reduce hard-easy effects and overconfidence bias through different educational strategy between student groups.

Key words: Overconfidence bias, hard-easy effect, biology concept, cognitive bias, self-monitoring



OPEN ACCESS

Brain, Digital, & Learning
2021, Vol. 11, No. 2, 307-319.
<https://doi.org/10.31216/BDL20210020>

Received: March 24, 2021

Revised: April 9, 2021

Accepted: May 21, 2021

© 2021. Institute of Brain based Education,
Korea National University of Education



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

학습자가 자신의 과제(task)에 대한 정확한 판단 또는 보정(calibration)을 하는 것은 자기조절 학습의 중요한 요소이며, 학문적 성취와 직접적인 관계가 있는 것으로 알려졌다 (Bol et al., 2010; Koku & Qureshi, 2004; Ots, 2013; Stone, 2000). 학생이 자신의 과제에 대한 이해 정도와 수행한 내용을 정확하게 평가하는 능력은 학습과정의 각 단계에서 매우 중요하다. 자기 학습에 대한 지속적인 모니터링을 통하여 반성적인 사고와 자기조절 능력을 함양시키는 것은 교육과정에서 강조하는 역량이기도 하다(MOE, 2015). 학생들은 자신의 목표를 향해 노력함에 따라, 그 노력의 성과를 모니터링하면서 자신의 전략과 학습에 대한 자원 배분을 조정하게 된다 (Hwang & Choi, 2020, Nelson, 1996; Pintrich, 2004; Winne, 2001). 만약 학습의 노력과 성과가 적절한 수준에서 수행되지 않는다고 판단하는 학생들은 통제력을 발휘하여 상황을 바로잡으려고 시도할 것이다 (Winne, 1995). 그러나, 모든 학습자가 이러한 학습모니터링에 성공적인 것은 아니다.

학습자가 자신이 과제를 수행하거나 성과를 낼 수 있다는 기대 정도와 자신감을 자기효능감(self-efficacy)이라고 한다. 긍정적인 자기효능감은 학습에 대한 동기와 의욕을 높이는데 효과적일 수 있지만, 지나치게 부풀린 자기효능감은 너무 높은 목표를 설정하게 되어 좌절과 포기를 동반하는 실패를 초래할 수 있다 (Bandura, 1986; Winne, 2004). 자신의 의사결정과 판단에 관한 연구를 보면 인간은 의사결정을 수행할 때 반성적이고 객관적인 판단보다는 직관적인 방법에 의존한다고 하였다 (Kahneman et al., 1982). 이러한 인간의 사고습관을 심리학자들은 인지편향(cognitive bias)이라고 정의하였다 (Hanselton et al., 2005). 이러한 인간의 사고 습관 중 이 연구에서는 합리적 의사결정과 자기 반성적 학습을 방해하는 것으로 알려진 과신편향(overconfidence bias)에 주목하고자 한다. 과신편향은 자신이 가지고 있는 능력보다 더 우수한 능력을 가지고 있다고 믿거나, 자신이 하는 방법이 틀렸을 수 있음에도 그것을 인식하지 못하고 옳다고 판단하는 성향으로 자신의 능력이나 판단을 객관적으로 점검하지 못하게 한다(Ha, 2017; Winman et al., 2004).

과신은 잘 입증된 인지편향 중 하나로, Weinstein (1980)이 학회 논문에서 처음 논의한 바에 따르면 학생들은 미래의 삶에 대해 지나치게 낙관적인 기대를 하고 있으며, 동료보다 더 높은 성과와 등급을 기대한다고 하였다. 더욱이 사람들은 다른 사람의 능력과 비교할 때 자신의 능력에 대해 지나치게 낙관적인 경향이 있다고 지적했다 (Taylor & Brown, 1988; Weinstein, 1980). Svenson (1981)은 대다수의 자동차 운전자가 자신이 다른 운전자보다 더 숙련되고 자신에게는 위험 부담이 적다고 생각한다는 것을 보여줬다. Wüst & Beck (2018)은 학습자의 과신은 교실에서 다음과 같은 몇 가지 결과를 초래할 수 있다고 정리하였다. 첫째, 학생이 자신의 능력에 대하여 잘 알고 있다면 시간과 자원을 더 효율적으로 할당하여 자신의 성적을 가장 많이 향상시킬 수 있도록 학습할 수 있다. 그러나, 과신은 학습에 대한 게으름과 집중력 부족으로 이어질 수 있다. 둘째, 실제 성적이 학생의 기대치를 충족하지 못할 때 성적에 대한 실망은 그 기대에 미치지 못한 과목에 대한 동기와 관심을 감소시킬 수 있다. 따라서 교실에서 과신의 결정 요인을 이해하는 것은 교육자와 학습자 모두에게 중요하다.

학생들을 대상으로 한 과신연구는 대부분의 학생들이 미래의 성과에 대해 지나치게 낙관적인 시각을 가지고 있다는 것을 발견했다. Nowell & Alston (2007)은 연구에 참여한 학생들의 58%가 자신의 성적을 정확하게 예측하였고, 그 그룹의 3분의 1은 자신이 받은 결과보다 더 나은 성적을 예상하였다고 밝혔다. 학생들 중 9%만이 자신감이 부족하였는데, 이들은 나중에 예상보다 더 좋은 성적을 받은 것을 확인하였다. 또한, 낮은 성

취도를 가진 학생들이 높은 성적을 받은 학생들보다 자신의 결과에 대해 더 낙관적이었으며, 남학생이 여학생보다 더 높은 수준의 과신을 보이는 것을 발견했다. Grimes (2010)는 거시 경제학 수업에서 나이와 전반적인 학업 성적이 감소하면 과신 정도가 증가하는 것을 발견하였다. 학생들은 특히, 남학생들이 여학생보다 많이, 학기의 모든 지점에서 자신의 성적을 과대평가 하였으나 학기가 진행됨에 따라 정확도가 향상되는 것이 발견되었다 (Beyer, 1999).

과신의 결정 요인은 다양하게 존재한다. Oskamp (1965)는 비록 그 정보가 도움이 되지 않더라도 더 많은 정보를 가졌다고 생각할수록 과신을 증가시킨다는 사실을 발견했으며, Russo & Schoemaker (1992)는 전문가들 사이에서도 과신이 존재하지만, 관련 주제에 대한 전문성이 증가함에 따라 그 정도는 감소함을 보여주었다. 반면 Glaser et al. (2005)는 주식 거래자들이 금융에 관한 질문에 답할 때 일반 사람들보다 지나치게 낙관적인 것을 발견하였다. 즉, 전문성이 항상 과신을 줄이는 것은 아니라는 것이다.

많은 연구에 따르면 자신의 결정에 대한 과대평가의 정도는 문화, 성별 및 작업 유형에 따라 영향을 받는다. Lichtenstein et al. (1977)는 질문에 올바른 답을 하는 자신의 능력에 대한 자신감이 더 쉬운 질문에 대해 낮은 경향이, 그리고 더 어려운 질문에 대해서는 높은 경향이 있음을 보여주었다. 이러한 경향은 또 다른 인지편향으로 "어려움-쉬움 효과(hard-easy effect)" 라고 한다 (Bordley et al., 2014). Larricket et al. (2007)은 사람들이 일반적으로 자신의 수행능력이나 답변에 대해 평균보다 더 높을 것이라고 인식하지만, 이 어려움-쉬움효과는 난이도가 다른 상황에서 역전될 수 있다는 것을 보여주었다. 즉, 어려움-쉬움효과란, 사람들이 보통 쉬운 문제나 과제에 대해서는 낮은 자신감을, 어려운 문제나 과제에 대해서는 오히려 높은 자신감을 드러내는 현상을 말한다 (Larricket et al., 2007). 저자들은 사람들이 특정 조건에서 어려운 문제에 대해서 과신감을 보여주는 반면, 쉬운 문제에 대한 자신감은 떨어지지만 사람들은 여전히 자신의 수행능력 또는 점수를 평균이상으로 기대한다고 하였다.

응답자의 자신감 정도와 실제 정답의 비율 차이는 한 개인의 과신에 대한 좋은 척도를 제공한다. 이 연구에서는 고등학교 학생들이 생물개념 확인 평가에서 갖는 개념 이해도를 측정함과 동시에 학생들의 과신감 정도를 확인하고자 하였다. Ha et al. (2012)은 과학개념을 평가하는 과정에서 학생들의 점수는 지식의 수준으로, 자신의 평가 점수에 대한 판단은 자기지식에 대한 확신이라고 정의하였다. 개념평가 문항에 대한 확신 정도를 스스로 판단하게 하는 척도를 통해서 비교적 간단하게 과신편향 정도를 검사할 수 있으며, 그 방법에 대한 타당도 및 신뢰도가 높다면 교육에서의 활용가치도 높을 것이다. 이 연구에서는 생명과학 개념평가 문항에 대한 학생들의 과신 인지 편향과 함께 문항의 난이도에 대한 학생들의 확신도의 차이를 보여주는 "어려움-쉬움 효과(hard-easy effect)"의 경향을 조사하고, 학생들이 갖는 유형을 그룹화 분석을 통해 확인해 보고자 한다.

효과적인 지식모니터링 기법은 학생의 답변에 대하여 개인이 어떤 항목의 의미를 알고 무엇을 알지 못하는지 성공적으로 식별할 수 있어야 한다. 학생들의 답변에 대한 신뢰도 판단과 정확성을 관련시키는 방법으로 이분법적 측정법이 사용될 수 있는데 민감도(sensitivity)와 특이도(specificity)라고 불리는 확증적 분석방법이 있다. 역학에서 민감도와 특이도는 임상 테스트를 평가하는 데 사용된다. 민감도는 질병이 있는 환자를 정확하게 식별하는 검사의 능력을 나타내며 특이도는 이러한 질병을 갖지 않는 환자를 정확하게 식별하는 검사 능력이다. Händel et al. (2020)은 민감도는 정확하게 감지된 정답의 상대적 빈도이고, 특이도는 정확하게 감지된 오답의 빈도라고 정리하였다. 이 연구에서는 민감도(sensitivity)와 특이도(specificity) 측정법을 활용하여 학생들이 정답에 대한 자신감과 실제 정답 비율을 확인하고 생물개념평가에서 나타나는 어려움-쉬움 효과

(hard-easy effect)의 경향을 확인하고자 한다. 지식모니터링 평가도구로서 민감도와 특이도는 Table 1과 같이 정리될 수 있다 (Rutherford, 2017; Schraw et al., 2013). 민감도와 특이도는 서로 보완이 되며 민감도는 올바르게 예측한 정답에 대한 비율을, 특이도는 올바르게 예측한 오답에 대한 비율을 나타낸다. 따라서 학생은 높은 민감도와 높은 특이도를 동시에 가질 수 있다.

Table 1. sensitivity and specificity

		Test Result	
		Correct	Incorrect
Expectancy	High Confidence	Correct/High Confidence (CH)	Incorrect/High Confidence (IH)
	Low Confidence	Correct/Low Confidence (CL)	Incorrect/Low Confidence (IL)

* sensitivity = $\frac{\sum CH}{\sum CH + \sum CL}$: Relative frequency of correctly detected correct answers

** specificity = $\frac{\sum IL}{\sum IH + \sum IL}$: Relative frequency of correctly detected incorrect answers

Methodology

Participants

이 연구에 참여한 학생들은 강원도 소재 일반계 A고등학교와 경기도 소재 일반계 B고등학교 1~2학년 학생 총 361명이었다. A고등학교에서는 남학생 110명(1학년 50명, 2학년 60명), 여학생 102명(1학년 47명, 2학년 55명)이 연구에 참여하였으며, B고등학교에서는 남학생 97명(1학년 51명, 2학년 46명), 여학생 62명(1학년 45명, 2학년 17명)이 연구에 참여하였다. 두 학교 모두 2학년들은 자연계열 과학탐구를 선택한 학생이었으며, 학생들은 2015개정 교육과정 중 공통교육과정인 통합과학의 개념을 모두 학습한 2학기 학기말에 본 평가에 임하였다.

Method

본 연구에 사용된 생물개념평가 검사지에 삽입된 문항은 그 타당성을 확보하기 위하여 미국과학진흥회 (Association for the Advancement of Science, AAAS)에 게시된 문항 중 생명관련 문항에서 발췌하였다(<http://assessment.aaas.org> 참조). 선정된 문항은 총 13개이며 생명시스템과 관련된 6문항, 생물다양성과 유지에 관한 4문항, 그리고 생태계와 환경에 대한 3문항으로 구성되어있다. 모든 문항은 선행연구들을 통해 한국어로 번역되었고 문항의 신뢰도(Cronbach's alpha값이 모두 0.6이상으로 신뢰할 수 있음)와 타당도(Rasch분석에서 MNSQ값이 모두 0.5~1.5 사이로 타당함)가 검증된 4지 선다형 문항들이다. 연구에 참여한 학생들이 선다형 문항에 응답한 후 자신의 답에 대한 확신 정도를 표시할 수 있도록 추가적인 문항을 제시하였는데, '확신이 없다'부터 '많이 확신한다'까지 1-5리커트 척도를 기반으로 문항을 구성하였다.

Data Analysis

생물개념에 대한 학생들의 지식 및 확신 수준에 따라 응답 유형을 나누기 위하여 먼저 검사지를 통해 제시한 13개 문항 중 정답 문항수의 평균을 산출하여 학생들의 지식수준을 확인했다. 확신 수준은 리커트척도 1~5점의 평균이다. 지식 점수와 자신에 답에 대한 확신 점수를 활용하여 높은 확신 정답, 낮은 확신 정답, 낮은 확신 오답, 높은 확신 오답의 4종으로 구성하였다. 정답이면서 확신 점수가 4~5점일 때 높은 확신 정답, 정

답이면서 확신 점수가 1~2일 때 낮은 확신 정답, 오답이면서 확신 점수가 1~2점일 때 낮은 확신 오답, 오답이면서 확신 점수가 4~5점일 때 높은 확신 오답이다. 확신점수 3점은 보통 확신 수준으로 응답 학생의 확신정도가 명확하게 드러나지 않아 본 연구 분석에서는 제외하였다. 이렇게 새롭게 구성된 점수를 바탕으로 평균을 내어 각 항목의 점수로 활용하였다. 생물개념에 대한 학생들의 지식모니터링 평가도구에서의 민감도(높은 확신 정답 비율/(높은 확신 정답+낮은 확신 정답))와 특이도(낮은 확신 오답/(높은 확신 오답+낮은 확신 오답))를 분석하였다.

본 연구에서는 학생들의 성별(남녀), 학년(1~2학년)에 따라 지식과 확신이 어떤 형태로 나타나는지 확인하기 위해 이원분산분석(two-way ANOVA)을 사용하였다. 자료의 내적 일관성을 파악하기 위하여 Cronbach's alpha를 이용하였고 Fisher (2007)가 제시한 기준에 따라 신뢰도가 0.67 이상일 때 적합하다고 판단하였다. Cronbach's alpha는 개념 평가가 0.725, 확신 수준평가가 0.939, 높은 확신 정답 0.847, 낮은 확신 정답 0.758, 낮은 확신 오답 0.811, 높은 확신 오답 0.790으로 모두 만족스러웠다. 학생들의 지식 및 확신 형태의 유사성을 파악하고 민감도와 특이도에 따라 최적의 집단을 제시하기 위하여 R패키지의 MCLUST를 사용하여 군집분석을 시행하였으며, 최적 집단수는 Bayesian Information Criterion (BIC) 값을 근거로 엘보우기법(elbow method)을 사용하여 판단하였다.

Results and Discussions

Fig.1은 개념점수와 확신점수가 성별과 학년에 따라 어떠한 특징을 보이는지를 나타내고 있다. 1학년 남학생은 2학년 남학생과 비교해 개념점수의 평균이 낮았으며(남학생 1학년 $M=0.617$, $SD=0.263$; 남학생 2학년 $M=0.651$, $SD=0.193$), 1학년 여학생도 2학년 여학생에 비하여 평균적으로 개념점수가 낮았다(여학생 1학년 $M=0.674$, $SD=0.210$; 여학생 2학년 $M=0.723$, $SD=0.177$). 하지만 전체적인 통계치로 보면 성별은 유의미한 차이를 보였으나($F[1, 316]=6.954$, $p=0.009$, $PES=0.022$), 학년급은 유의미하지 않았다($F[1, 316]=2.843$, $p=0.093$, $PES=0.009$). 성별과 학년급별 교호작용효과도 유의미하지 않았다($F[1, 316]=0.106$, $p=0.745$, $PES=0.000$).

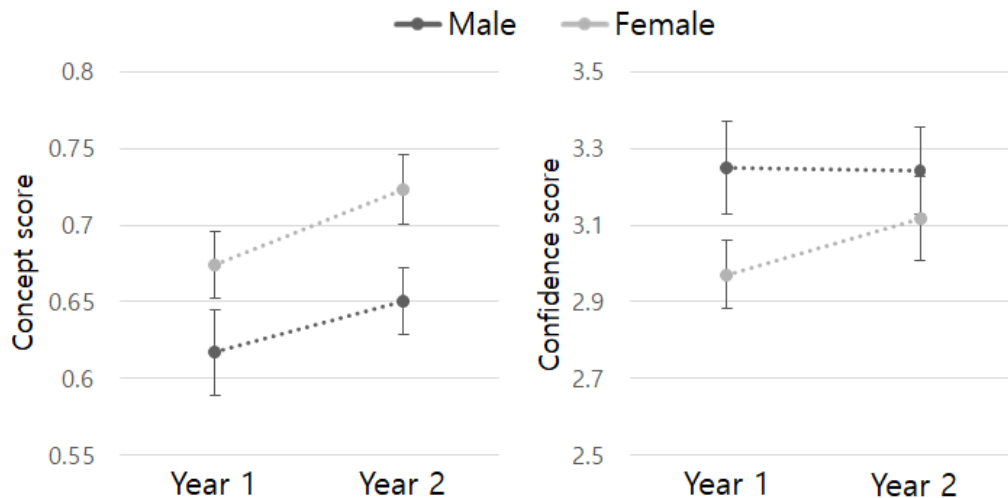


Fig. 1. Conceptual scores and confidence scores between genders and years

확신 측면에서 살펴보면, 남학생들은 학년급 간 확신 점수 차이가 크지 않았으나(남학생 1학년 $M=3.250$, $SD=1.154$; 남학생 2학년 $M=3.242$, $SD=1.005$), 여학생들은 1학년보다 2학년에서 높은 확신 점수를 보였다(여학생 1학년 $M=2.971$, $SD=0.848$; 여학생 2학년 $M=3.117$, $SD=0.842$). 전체적으로 볼 때 성별, 학년급에서 모두 통계적으로는 유의미한 차이를 보이지 않았으며(성별 $F[1, 316]=3.295$, $p=0.070$, $PES=0.010$; 학년급 $F[1, 316]=0.386$, $p=0.535$, $PES=0.001$) 성별과 학년급별 교호작용효과도 유의미하지 않았다($F[1, 316]=0.481$, $p=0.488$, $PES=0.002$)

여학생들의 확신 수준이 학년 진급에 따라 향상되고 있는 것으로 보아, 여학생 집단에서는 정답 여부와 관련 없이 지식수준이 향상됨에 따라 자연스레 지식에 대한 확신 또한 함께 증가한 것으로 예상해볼 수 있다. 반면 남학생들은 여학생에 비하여 낮은 지식수준임에도 불구하고 높은 수준의 확신을 보였는데, 여학생들에 비해 남학생들에서 과신편향(overconfidence bias)이 더 자주 발생한다는 선행연구들과 같은 결과로 보인다(Keller & Molix, 2008; Pajares & Graham, 1999; Pallier, 2003; Rachmatullah & Ha, 2019; Stankov & Lee, 2008; Wood & Eagly, 2012).

성별 및 학년별 간에 개념과 확신의 정도가 어떠한 유형으로 나타나고 있는지 Fig. 2를 통해 살펴볼 수 있다. 정답을 선택한 학생들의 확신정도를 비교 분석한 결과, 대체적으로 1학년 남학생들이 2학년 남학생들에 비해 높은 확신을 가지고 정답을 선택한 것으로 나타났으며(남학생 1학년 $M=0.370$, $SD=0.301$; 남학생 2학년 $M=0.352$, $SD=0.255$), 여학생의 경우에는 남학생과 같은 양상을 보였으나 학년간 큰 차이가 발견되지는 않았다(여학생 1학년 $M=0.348$, $SD=0.263$; 여학생 2학년 $M=0.340$, $SD=0.263$). 전체적으로는 성별과 학년 모두에서 유의미한 통계적 차이를 보이지 않았으며(성별 $F[1, 314]=0.294$, $p=0.588$, $PES=0.001$; 학년 $F[1, 314]=0.176$, $p=0.675$, $PES=0.001$) 성별과 학년의 교호작용효과도 유의미하지 않았다($F[1, 314]=0.032$, $p=0.858$, $PES=0.000$). 정답을 선택했음에도 낮은 확신을 보인 경우, 1학년 남학생들에 비해 2학년 남학생들에게서 낮은 확신을 나타내는 경우가 평균적으로 약간 높게 나타났으며(남학생 1학년 $M=0.132$, $SD=0.163$; 남학생 2학년 $M=0.159$, $SD=0.203$), 여학생들의 경우에도 남학생들과 동일한 양상을 보였으나 그 폭이 남학생들에 비해 비교적 크게 나타났다(여학생 1학년 $M=0.166$, $SD=0.168$; 여학생 2학년 $M=0.176$, $SD=0.217$). 통계적으로는 전체 성별과 학년 모두에서 유의미한 차이는 나타나지 않았으며(성별 $F[1, 314]=1.462$, $p=0.228$, $PES=0.005$; 학년 $F[1, 314]=0.803$, $p=0.371$, $PES=0.003$), 성별 및 학년 교호작용효과도 유의미하지 않았다($F[1, 314]=0.155$, $p=0.694$, $PES=0.000$).

학생들의 확신 정도와 오답 선택 정도를 비교한 결과, 남학생 집단(남학생 1학년 $M=0.196$, $SD=0.243$; 남학생 2학년 $M=0.141$, $SD=0.179$)과 여학생 집단(여학생 1학년 $M=0.198$, $SD=0.189$; 여학생 2학년 $M=0.144$, $SD=0.147$)에서 학년별로 비슷한 모습을 보였는데 두 집단 모두 1학년 학생들이 2학년 학생들에 비해 오답을 선택하며 낮은 확신을 갖는 것으로 나타났다. 전체적으로 성별에서는 유의미하지 않았으나($F[1, 314]=0.010$, $p=0.922$, $PES=0.000$) 학년급에서는 유의미한 차이를 나타냈으며($F[1, 314]=5.936$, $p=0.015$, $PES=0.019$) 성별과 학년급의 교호작용효과는 유의미하지 않았다(교호작용 $F[1, 314]=0.002$, $p=0.968$, $PES=0.000$).

한편, 1학년 남학생들보다 2학년 남학생들이 평균적으로 높은 확신을 가지고 오답을 선택하는 경향이 두드러졌으며(남학생 1학년 $M=0.098$, $SD=0.160$; 남학생 2학년 $M=0.124$, $SD=0.166$) 여학생들 또한 1학년에 비해 2학년 학생들이 평균적으로 비교적 높은 확신을 가지고 오답을 선택하는 것으로 나타났다(여학생 1학년 $M=0.059$, $SD=0.123$; 여학생 2학년 $M=0.073$, $SD=0.134$). 전체적으로 봤을 때 성별에서 유의미한 차이가 나타

났으나($F[1, 314]=7.126, p=0.008, PES=0.022$) 학년에서는 통계적으로 유의미한 차이는 아닌 것으로 확인되었다($F[1, 314]=1.422, p=0.234, PES=0.005$). 성별과 학년의 교호작용효과도 유의미하지 않은 것으로 나타났다($F[1, 314]=0.133, p=0.716, PES=0.000$).

1학년에서 2학년으로 학년이 올라가며 생물개념에 대해 학습하는 양이 증가하고 지식에 대한 전문성이 높아질 것으로 예상할 수 있다. 하지만 본 연구결과, 학년이 올라감에 따라 오답을 선택할 때 낮은 확신을 보이는 경우보다, 높은 확신을 보인 경우가 유의미하게 높게 나타났다. 이는 Glaser et al. (2005)의 연구결과와 같이 전문성의 증가가 과신의 감소를 의미하지는 않는다고 생각해볼 수 있다.

오답을 선택할 때 여학생 집단보다 남학생집단에서 유의미하게 높은 자신감을 보인 것은 앞서 논의한 것과 같이 여학생들보다 남학생들에게 과신편향(overconfidence bias) 및 더닝크루거효과(Dunning-Kruger effect)가 비교적 더 작용하고 있는 결과라 추측해볼 수 있다.

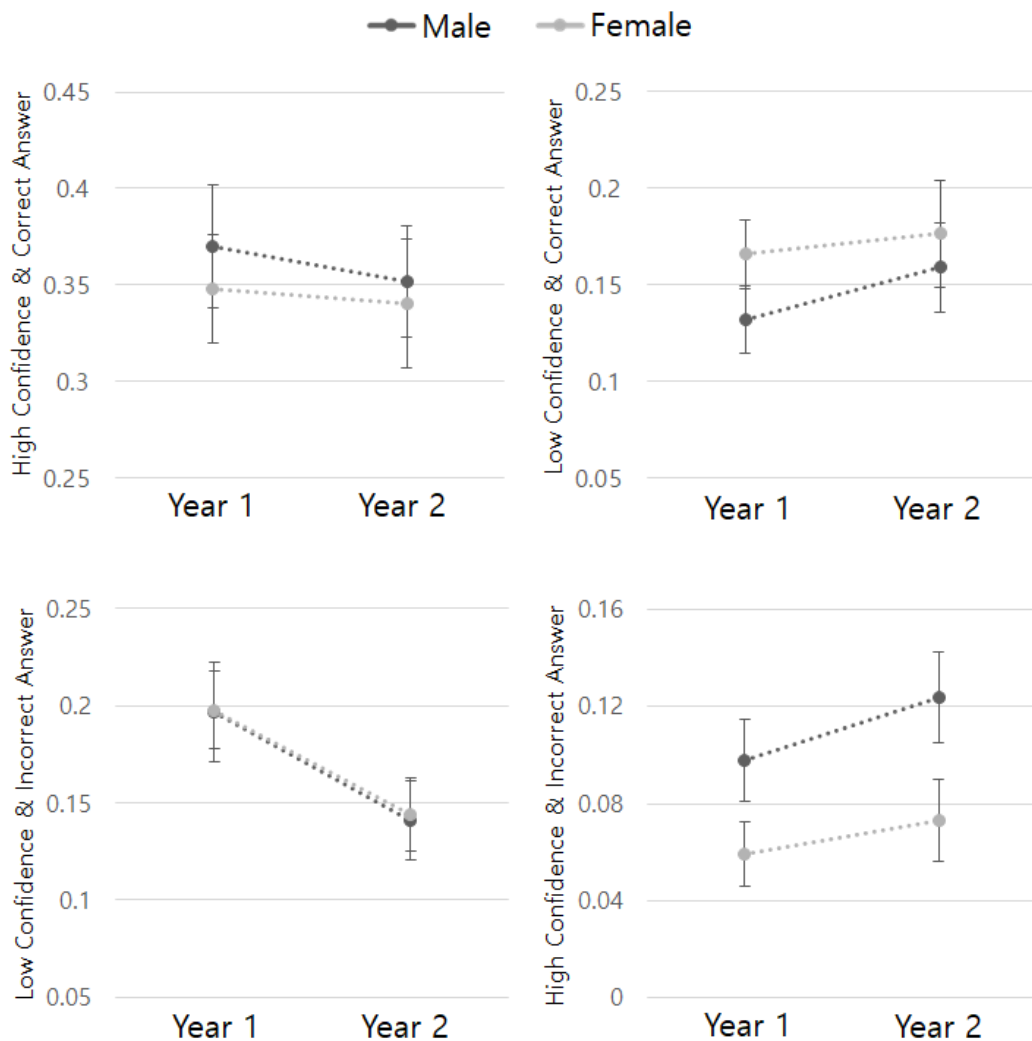


Fig. 2 Conceptual-confidence type between gender and year

Fig. 3은 성별과 학년별로 민감도와 특이도에 어떤 특징을 보이는지를 나타내고 있다. 정답을 선택한 학생들 중 높은 자신감을 가지고 정확하게 정답을 예측한 학생들의 비율을 나타내는 민감도는 1학년 남학생들에 비해 2학년 남학생들에게서 평균적으로 높게 나타났으나(남학생 1학년 $M=0.672$, $SD=0.365$; 남학생 2학년 $M=0.687$, $SD=0.360$), 여학생 집단에서는 학년 간 비슷한 수치를 보였다(여학생 1학년 $M=0.628$, $SD=0.329$; 여학생 2학년 $M=0.624$, $SD=0.382$). 전체적으로 볼 때, 성별과 학년에 따른 차이는 유의미하지 않았으며(성별 $F[1, 304]=1.686$, $p=0.195$, $PES=0.006$; 학년 $F[1, 304]=0.019$, $p=0.891$, $PES=0.000$) 성별과 학년의 교호작용효과 또한 유의미하지 않았다($F[1, 304]=0.055$, $p=0.814$, $p=0.000$)

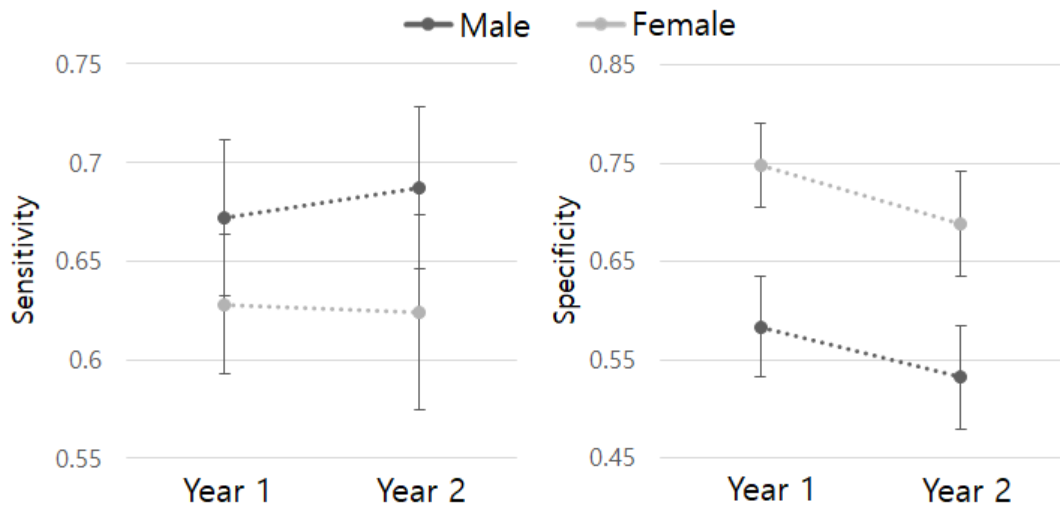


Fig. 3. Sensitivity and specificity by gender and grade

오답을 선택한 학생 중 자신의 답이 오답일거라 정확하게 예측한 학생들의 비율을 나타내는 특이도는 2학년 남학생들에 비해 1학년 남학생들에게서 평균적으로 높게 나타났으며(남학생 1학년 $M=0.584$, $SD=0.455$; 남학생 2학년 $M=0.532$, $SD=0.443$) 여학생들에게서도 1학년이 2학년에 비해 높은 것으로 나타났으나(여학생 1학년 $M=0.748$, $SD=0.363$; 여학생 2학년 $M=0.688$, $SD=0.390$) 유의미한 정도의 차이가 발견되지는 않았다. 전체적으로 성별에서 유의미한 결과가 나타났으나($F[1, 275]=10.039$, $p=0.002$, $PES=0.035$) 학년에는 유의미하지 않았으며($F[1, 275]=1.218$, $p=0.271$, $PES=0.004$) 성별과 학년의 교호작용효과도 유의미하지 않았다($F[1, 275]=0.007$, $p=0.935$, $PES=0.000$).

위의 결과를 살펴보면 여학생들이 남학생들보다 높은 특이도를 보이고 있었다. 다시 말해, 여학생들이 남학생들에 비해 자신이 선택한 답안이 오답일 가능성을 비교적 정확하게 예측하고 있었다. 반면, 남학생들은 여학생들에 비해 자신의 답안이 오답일 가능성을 잘 예측하지 못함을 나타내고 있으며 이는 남학생들이 여학생들과 비교하여 더 높은 수준의 과신 편향을 보인다고 제시한 Nowell & Alston (2007)의 선행연구와 일치한다.

학생들의 지식 개념 및 확신 형태의 유사도에 근거하여 모형기반 군집분석을 시행한 결과, 총 5개 집단이 최적 집단 수로 나타났으며, 이는 Fig. 4와 같다. 이 분석에는 총 361명의 데이터에서 군집분석에 유효한 338명의 데이터만을 사용하였다. 집단1의 학생들은 비교적 높은 확신을 가지고 정답을 선택하며 오답의 경우 높은 확신을 거의 보이지 않은 유형으로 총 88명(26.0%)의 학생들이 속하였다. 집단2에 속하는 학생은 29명(8.6%)으로, 집단1의 학생들과 유사하게 높은 확신을 가지고 정답을 선택하지만 오답을 선택할 때 또한 비교적 높은 확신을 보이는 것으로 나타났다. 집단3의 학생들은 87명(25.7%)이며, 전반적으로 정답과 오답 선택 여부와 상관없이 모든 문항에 대해 낮은 확신을 가지고 있었다. 반면, 집단4의 학생들(70명, 20.7%)은 모든 문항에 대해 상당히 높은 확신을 가진 것으로 나타났으며 낮은 확신을 거의 보이지 않았다. 마지막으로 집단5(64명, 18.9%)에 속하는 학생들은 집단2의 학생들과 비교적 유사한 형태를 보였으나, 집단2의 학생들과는 다르게 오답을 선택할 때 높은 확신을 보이는 경우가 적은 것을 확인할 수 있었다.

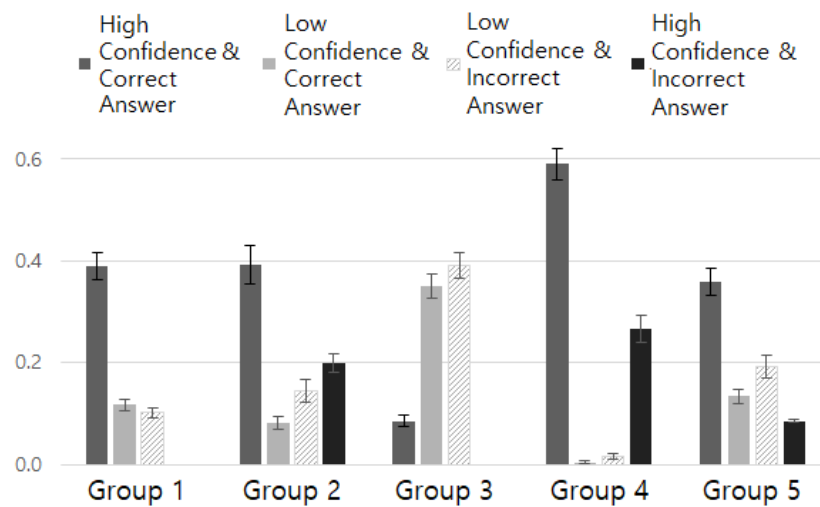


Fig. 4. Cluster model based on similarity of knowledge and confidence

Bordley et al. (2014)이 말한 어려움-쉬움 효과(hard-easy effect)의 두 가지 측면 즉, 정답을 선택했음에도 불구하고 자신의 답안에 대한 확신이 낮으면서, 오답을 선택했음에도 높은 자신감을 보이는 경우가 동시에 두드러지게 나타나는 집단은 본 연구에서 발견되지 않았다. 하지만 집단 2와 5에서 일부 그러한 현상을 발견할 수 있었는데, 이들 집단의 학생들은 정답을 선택하면서 확신감이 낮고, 오답을 선택하면서 비교적 높은 확신감을 갖는 양상을 보였다. 이러한 경향을 보인 학생들이 전체 연구 참여자 중 27% 이상인 것으로 나타났다. 이와 반대로 두 측면 중 한 측면이 눈에 띄게 나타나는 집단들(집단 3과 4)이 확인되었다. 집단3의 경우 전체의 25.7% 학생이 속해있으며 이들은 정답에 대한 낮은 확신을 보이고 있는데 이들은 자신의 지식과 능력을 과소평가하는 경향이 있는 학생들로 구성된 집단이다. 반면, 집단4의 경우 오답을 선택하면서도 높은 자신감을 보이는 학생들로 구성되어있는데 이들은 전체 참여자의 20.7%를 차지하며, 자신이 해결할 수 없는 문항을 마주하면서도 본인의 능력을 과신하는 경향이 있는 것으로 확인되었다.

민감도와 특이도를 기준으로 학생들의 최적집단 수를 분석한 결과 총 4개의 집단으로 구성될 수 있음을 확인하였으며 Fig. 5의 그래프와 같이 나타났다. 이 군집분석에는 총 361명의 데이터 중 분석에 유효한 295명의 데이터가 사용되었으며, 4개의 집단은 각각 명확하게 구별되는 특징을 가지고 있는 것으로 나타났다. 먼저 집단1(16.0%)의 학생들에게서는 민감도가 특이도보다 높게 나타났다. 즉, 정답을 선택한 경우, 정답을 정확하게 예측한 정도가 굉장히 높으며 이에 비해 오답을 예측하는 정도는 비교적 낮았다. 집단2(25.4%)의 학생들의 민감도와 특이도는 매우 바람직한 모습을 보이고 있다. 이 학생들은 정·오답에 대해 정확한 예측이 가능하며 이는 자신이 알고 모르는 지식에 대해 분명하게 구별하고 있다는 것을 의미한다. 집단3의 학생들과 집단4의 학생들은 서로 완전히 다른 양상을 보였다. 집단3(29.2%)의 경우 민감도는 매우 높으나 특이도는 거의 없는 형태를 보였는데, 이는 자신의 답안이 정답일 거라 예측하는 정도는 매우 높지만 오답에 대한 예측 정도는 매우 낮은 형태였다. 즉, 정·오답 모두에 높은 확신을 보이는 집단이라고 분류할 수 있다. 반대로 집단 4(29.5%)의 학생들은 민감도가 극히 낮은 반면 특이도는 매우 높은 형태를 보였는데, 이는 정답에 대한 예측이 매우 낮고 오답에 대한 예측이 극히 높은 형태로, 본인의 답안에 대해 전체적으로 확신이 없는 상태로 보인다.

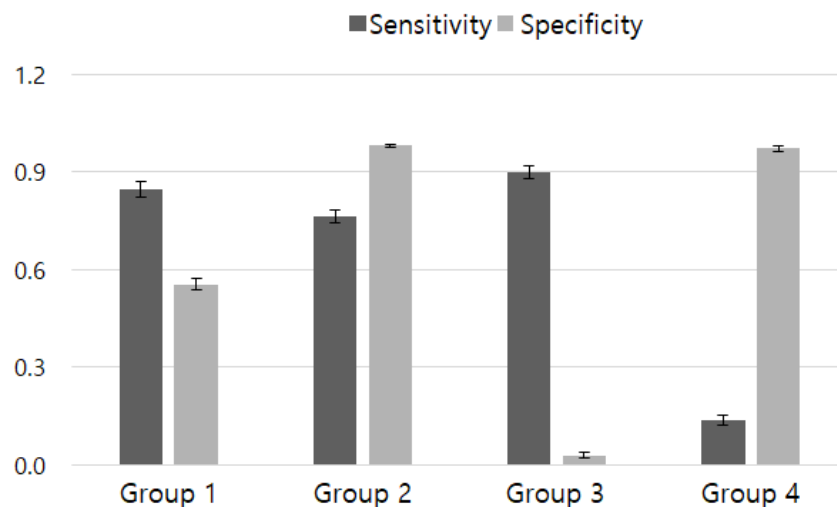


Fig. 5. Cluster model based on similarity of sensitivity and specificity

학생들의 생물개념 지식 및 확신에 대한 민감도와 특이도를 기준으로 분류한 4개의 집단은 직전 논의에서와 같이 교사들이 각 집단별로 다른 교육처치가 필요함을 제안하고 있다. 특별히 모든 문항에서 낮은 자신감을 보이는 집단3의 학생들과 모든 문항에서 높은 확신도 즉, 과신편향을 보이는 집단4의 학생들이 전체 연구 참여자의 58% 이상을 차지하고 있음이 확인되었다. 이 결과를 토대로, 교사는 집단 3의 학생들에게는 생물개념과 지식을 정확하게 습득할 수 있도록 교육하고 집단 4의 학생들에게는 과신편향을 인지하고 감소시킬 수 있도록 교육하는 등 집단 간 다른 교육방법을 적용해야 한다. 또한 일반적으로 이러한 편향적 사고는 학생들의 합리적 의사결정을 방해할 수 있기에, 적절한 교육을 통해 학생들이 스스로 편향적 사고를 인지하고 줄일 수 있도록 도와야 함을 시사하고 있다(Rachmatullah & Ha, 2019).

Conclusions and Educational Implications

이 연구에서는 생물개념평가 검사지를 통해 학생들의 생물영역 지식수준 및 확신수준을 알아보고 지식과 확신 간 4가지 유형에서 성별 및 학년별간 어떠한 특징이 나타나는지, 그리고 민감도와 특이도에 있어 유의미한 차이가 나타나는지 살펴보고자 하였다. 연구결과를 종합하여 몇 가지 다음과 같은 결론을 내렸다.

첫째, 학년과 상관없이 여학생 집단보다 남학생 집단에서 과잉확신 편향이 두드러지게 나타났다. 학생들의 지식수준을 분석한 결과 남학생의 생명과학 개념지식 수준이 여학생들보다 유의미하게 낮은 것으로 확인되었으나 지식-확신 유형 분석 시, 높은 확신을 가지고 오답을 선택한 경우는 여학생들보다 남학생들 집단에서 유의미하게 높게 나타났다. 즉, 남학생들에게서 과신편향이 두드러지게 보인다는 것이다. 이는 남학생들이 여학생들에 비해 과신편향에 빠질 가능성이 높다는 선행연구들과 일치한 결과이다(Ha, 2019; Keller & Molix, 2008; Pajares & Graham, 1999; Pallier, 2003; Rachmatullah & Ha, 2019; Stankov & Lee, 2008; Wood & Eagly, 2012).

둘째, 지식 및 확신의 유사성에 근거하여 군집 모형 분석을 시행한 결과, 많은 수의 학생들에게서 어려움-쉬움 효과가 나타났다. 지식-확신의 형태를 기준으로 5개의 최적 집단이 산출되었는데 이 중 한 개의 집단(26.0%)을 제외한 모든 집단에서 어려움-쉬움 효과(hard-easy effect)가 일부 또는 두드러지게 나타나고 있었다. 또한, 민감도 및 특이도의 유사성에 근거한 군집 모형 분석을 시행한 결과, 과반수 이상(58.7%)의 학생들에게서 편향적 사고가 드러났다. 이 중 절반의 학생들은 모든 문항에서 확신의 정도가 매우 낮았으며, 나머지 절반의 학생들에게서는 모든 문항에서 과신의 경향이 나타났다. 이는 생물교육에 시사하는 바가 매우 크다. 어려움-쉬움 효과(hard-easy effect)는 일종의 편향적 사고로 학습자의 학습동기와 학습태도에 좋지 못한 영향을 줄 수 있다는 것이 선행연구들의 분석이다(Wüst & Beck, 2018). 따라서 교사는 각 집단을 구성하는 학생들의 특성에 맞게 교육적 처치를 적용하여 편향적 사고를 바로잡아야 할 필요성을 제시하고 있는 것이다.

이와 같은 연구의 결론들을 통해 생물교육에 대해 제언한다. 첫째, 자기주도학습에서 학생들이 자신의 학습상태와 개념의 이해정도를 스스로 파악하는 것이 중요하다. 안다고 느끼는 것과 실제로 아는 것을 구분할 수 있는 역량을 향상시킬 수 있도록 정답만 맞히는 교육이 아닌 자신의 확신정도를 표현할 수 있도록 하고, 평가 후 지식과 확신 간의 일치도를 피드백 해준다면 학생들이 스스로 편향적 사고를 인지하고 점차적으로 줄일 수 있을 것으로 예상된다. 둘째, 개별 학생들이 가지고 있는 역량에 따라 맞춤형 교육을 진행하는 것은 매우 중요한 일이나, 현실적으로 학생 개개인에게 맞춤형 교육을 시행하는 것에는 어려움이 따른다. 군집모형 분석을 통해 유사한 특징을 가진 학생들을 집단으로 분류하고 교사들이 각 집단 간 특성을 분석하여 적합한 학습법을 제안할 때 학생들이 보다 더 나은 교육적 성취를 이룰 수 있을 것으로 보인다.

References

- Bandura, A. (1986). *Social Foundation of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs N.J: Prentice-Hall.
- Beyer, S. (1999). Gender differences in the accuracy of grade expectancies and evaluations. *Sex Roles*, 41, 279-296.
- Bol, L., Riggs, R., Hacker, D. J., & Nunnery, J. (2010). The calibration accuracy of middle school students in math classes. *Journal of Research in Education*, 21, 81-96.

- Bordley, R., LiCalzi, M., & Tibiletti, L. (2014). A Target-based Foundation for the “Hard-easy Effect” Bias. Universit`a Ca’ Foscari Venezia, Italia, Working Paper.
- Francis, X. S., & Christopher, A. W. (2014). Knowledge Monitoring Calibration: Sensitivity and Specificity as Unique Cognitive Constructs. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36, 3078-3083.
- Glaser, M., Langer, T., & Weber, M. (2005). Overconfidence of professionals and lay men: Individual differences within and between tasks? Working Paper. Sonderforschungsbereich 504 Publications (No. 05-25). retrieved from <http://EconPapers.repec.org/RePEc:xrs:sfbmaa:05-25>
- Grimes, P. W. (2010). The overconfident principles of economics student: An examination of a metacognitive skill. *The Journal of Economic Education*, 33, 15-30.
- Ha, M., Haury, D. L., & Nehm, R. H. (2012). Feeling of certainty: Uncovering a missing link between knowledge and acceptance of evolution. *Journal of Research in Science Teaching*, 49, 95-121.
- Ha, M. (2017). Pre-service biology teachers’ mixed model of Darwinism and Lamarckism. *Brain, Digital, & Learning*, 7, 175-189.
- Ha, M. (2019). Biology teachers’ perception of value of biology learning. *Brain, Digital, & Learning*, 9, 1-8.
- Handel, M., Harder, B., & Dresel, M. (2020). Enhanced monitoring accuracy and test performance: Incremental effects of judgment training over and above repeated testing. *Learning and Instruction*, 65, 101245.
- Haselton, M.G., Nettle, D. & Andrews, P.W. (2005). The evolution of cognitive bias. In D. M. Buss (Ed.), *Handbook of Evolutionary Psychology* (pp. 724-746). Hoboken: Wiley.
- Hwang, S., & Choi, H., (2020). The influence of group presentations physics class enhanced self-directed learning on science-related attitudes and communication skills of meister high school students. *Brain, Digital, & Learning*, 10, 499-501.
- Kahneman D., Slovic P., & Tversky, A. (1982) *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Keller, J., & Molix, L. (2008). When women can’t do math: The interplay of self-construal, group identification, and stereotypic performance standards. *Journal of Experimental Social Psychology*, 44, 437-444.
- Koku, P. S., & Qureshi, A. A. (2004). Overconfidence and the performance of business students on examinations. *Journal of Education for Business*, 79, 217-224.
- Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you’re better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes*, 102, 76-94.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In H. Jungermann & G. Zeeuw (Eds.), *Decision Making and Change in Human Affairs*. Dordrecht, Netherlands: Reidel, 275-324.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, 51, 102-116.
- Nowell, C., & Alston, R. M. (2007). I thought I got an A!: Overconfidence across the economics curriculum. *The Journal of Economic Education*, 38, 131-142.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29, 261-265.
- Ots, A. (2013). Third graders' performance predictions: Calibration deflections and academic success. *European Journal of Psychology of Education*, 28, 223-237.
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24, 124-139.

- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles*, 48, 265–276.
- Pintrich, P. (2004). A conceptual framework for assessing motivation and self regulated learning in college students. *Educational Psychology Review*, 16, 385–407.
- Rachmatullah, A., & Ha, M. (2019). Examining high-school students' overconfidence bias in biology exam: a focus on the effects of country and gender. *International Journal of Science Education*, 41, 652–673.
- Russo, J. E., & Schoemaker, P. J. H. (1992). Managing overconfidence. *Sloan Management Review*, 33, 7–17.
- Rutherford, T. (2017). The measurement of calibration in real contexts. *Learning and Instruction*, 47, 33–42.
- Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, 24, 48–57.
- Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, 100, 961–976.
- Stone, N. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12, 437–475.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47, 143–148.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39, 806–820.
- Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: how to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1167–1175.
- Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, 30, 173–187.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Self-regulated Learning and Academic Achievement: Theoretical Perspectives* (2nd ed) (pp. 153–189). Mahwah, NJ: Erlbaum.
- Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research*, 41, 466–488.
- Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. *Advances in Experimental Social Psychology*, 46, 55–123.
- Wüst, K., & Beck, H. (2018). “I thought I did much better”—Overconfidence in university exams. *Decision Sciences Journal of Innovative Education*, 16, 310–333.

Authors Information

Hwang, Hyojeong: Kangwon National University, Doctoral student, First Author

ORCID: <https://orcid.org/0000-0001-8006-8950>

Ha, Minsu: Kangwon National University, Professor, Co-author

ORCID: <http://orcid.org/0000-0003-3087-3833>

Park, Eunju: Kangwon National University, Postdoctoral researcher, Corresponding Author

ORCID: <https://orcid.org/0000-0003-4675-4919>