

과제 수행 능력의 부족에 대한 몰이해: Kruger와 Dunning(1999)의 학업 분야에서의 반복 연구

박 상 희[†]

충북대학교 심리학과

Kruger와 Dunning(1999)은 특정 분야에서 과제 수행 능력이 부족한 사람일수록 자신의 수행을 실제보다 더 긍정적으로 추정함을 보였다. 이 연구에서는 학부 수업에서 진행한 자료 수집을 통해, 이 소위 Dunning-Kruger 효과를 현실의 학업 맥락에서 반복 검증하였다. 학생들은 기말에 치러진 시험 직후에 자신의 객관적인 수행을 추정하였다. 이 추정과 실제 점수를 비교한 결과 Dunning-Kruger 효과가 관찰되었다. 시험 준비에 들인 노력은 실제 점수를 증가시켰으므로 과대 추정을 감소시키는 간접 효과와 과대 추정을 증가시키는 직접 효과를 모두 나타냈다. Dunning-Kruger 효과가 지각된 어려움에 의해 나타난 인위물(artifact)이라는 대안 설명은 뒷받침되지 않았으며, 수행의 추정이 얼마나 구체적이거나 추상적인 수준에서 이루어지는지는 과대 추정 정도에서의 유의미한 차이로 이어지지 않았다. 종합하면 본 연구의 결과는 Dunning-Kruger 효과의 설명 중 '평균으로의 회귀' 설명(Krueger & Mueller, 2002)이나 '과제 난이도' 설명(Burson, Larrick, & Klayman, 2006)보다는 Kruger와 Dunning(1999)의 '메타 인지' 설명에 더 일치하는 결과를 보이고 있다. 다른 대안 설명의 가능성과 한계, 후속 연구에 대해 논의하였다.

주요어 : 수행, 메타인지, 자기평가, 과잉확신

[†] 교신저자: 박상희, 충북대학교 사회과학대학 심리학과, (361-763) 충북 청주시 흥덕구 내수동로 52
Tel: 043-261-2192, E-mail: sangheepark@chungbuk.ac.kr

사람들은 자신의 능력에 대해 얼마나 잘 이해하고 있는가? 2000년의 Ig Nobel 상을 수상한 연구에서 Kruger와 Dunning(1999)은 사람들이 흔히 주어진 과제에서의 자신의 상대적인 수행을 현실보다 더 긍정적으로 보는 경향이 있음을 밝혔다. 더 중요한 발견은 이러한 경향이 해당 과제에서의 능력이 낮은 사람일수록 더 심하게 나타난다는 것이었다.

전형적인 실험에서 참가자들은 주어진 과제를 수행한 뒤에 자신의 점수가 전체 참가자들 중에서 어느 정도 순위(백분위)에 위치할지(백분위)를 추정하였다. 그 결과 대체로 실제 수행이 낮을수록 자신의 수행에 대한 기대도 낮아지긴 했지만 그 기울기는 실제 수행보다 훨씬 완만하며, 수행이 아주 낮은 참가자들은 자신의 수행을 비현실적으로 높게 추정하였다. Kruger와 Dunning(1999)은 이 경향이 유머 감각, 논리적 추리, 문법 실력 등의 다양한 분야에서 나타남을 보였고, 이후 다른 연구자들이 다양한 참가자 집단과 과제 분야에 대해서 반복 검증하였다(예: 의대생들의 면접 기술; Parikh, McReelis, & Hodges, 2001). 한편 능력이 낮은 사람들의 과대 추정 경향과는 반대로 능력이 뛰어난 사람들은 오히려 자신의 수행을 과소 추정하는 경향을 나타냈다. 이런 두 가지 경향 중에 수행이 낮은 참가자들의 과대 추정은 그 효과의 크기가 더 클 뿐 아니라 현실적인 함의도 더 중요하므로 이것이 대부분의 연구의 주된 관심사가 되고 있다(Krueger & Mueller, 2002; Burson, Larrick, & Klayman, 2006 등). 대개 Dunning-Kruger 효과는 수행이 낮은 참가자들이 자신의 수행을 현실보다 높게 추정하는 현상을 가리키는 말로 사용되고 있지만, 이 현상에 관한 대부분의 연구들은 수행이 높은 참가자들의 과소 추정까지도 함께 설명하는 것을 목표로 하고 있다.

왜 이런 결과들이 나오는지에 대해 Kruger와 Dunning(1999)은 메타 인지(metacognition)의 측면에서 설명하였다. 즉 어떤 과제와 관련한 자신의 능력을 평가할 수 있으려면 그 과제에 대한

전문 지식이 있어야 하는데, 능력이 낮은 사람들은 이런 전문 지식이 없기 때문에 자신의 능력을 정확히 평가할 수 없으며 동시에 수행도 낮아지게 되는 것이다. 이런 결과를 Dunning 등(Dunning, Johnson, Ehrlinger, & Kruger, 2003)은 이중의 저주(double curse)라고 불렀다. 반면에 능력이 뛰어난 사람들이 자신의 수행을 과소 추정하는 이유에 대해서 Kruger와 Dunning은 그들이 남들을 과대평가해서 남들도 자신처럼 수행이 높을 거라고 생각하기 때문에 자신의 상대적인 수행을 실제만큼 높게 추정하지 않는 것이라고 설명했다. 즉 능력이 낮은 사람들은 자기 추정에 실패하는 반면 능력이 높은 사람들은 타인 추정에 실패한다는 설명이다. 정리하면 Dunning-Kruger 효과에 대해 Kruger와 Dunning(1999) 본인들은 동기적 설명을 배제한 순수한 메타 인지적 설명을 제시하고 있다. 실제로 Ehrlinger, Johnson, Banner, Dunning과 Kruger(2008)의 최근 연구에서는 100달러까지 걸면서 금전에 의한 동기화를 유도했지만 추정의 정확성은 증가하지 않았다.

다른 연구자들은 Dunning-Kruger 효과의 존재 자체는 부정하지 않지만, 그 효과의 설명을 위해 능력에 따른 메타 인지에서의 차이가 아닌 전혀 다른 메커니즘들을 동원하고 있다. 대안 설명들 중에 대표적인 것들을 들면, 먼저 Krueger와 Mueller(2002)는 단순히 평균 이상 효과(Better-Than-Average effect: Krueger, 1998)와 평균으로의 회귀 현상이 동시에 일어난 결과 나타나는 것이라고 주장했다. 즉 과제 수행의 추정이 신뢰롭지 못할수록 실제 능력과 추정된 능력의 상관은 낮아진다. 이는 실제 수행과 수행의 추정의 관계에 있어서 평균으로의 회귀로 이어지고, 실제로는 사람들이 어느 정도 자신의 수행을 추정할 수 있음에도 불구하고 과대 추정(능력이 낮은 사람들의 경우)과 과소 추정(능력이 높은 사람들의 경우)의 현상이 나타나게 되는 것이다. 또한 모든 사람들이 자신의 능력을 과대 평가하는 경향이 있는데, 능력이 낮은 사람

들에 있어서는 이 경향이 평균으로의 회귀와 합쳐져서 과대 추정이 극대화되고, 능력이 높은 사람들의 경우에는 평균으로의 회귀와 서로 상쇄 작용을 일으키기 때문에 결과적으로 약간의 과소 추정이 나타나는 것이라는 주장이다. 이 주장에 의할 경우 Dunning-Kruger 효과가 발생하기 위해서는 평균으로의 회귀와 평균 이상 효과가 모두 존재해야 하므로 둘 중 하나라도 발생하지 않는 조건에는 Dunning-Kruger 효과가 발생하지 않는다. 그런데 평균 이상 효과는 상대적인 수행에 대해서만 발생하는 것이므로(즉 자신이 평균보다 낫다고 생각하는 것이므로) 절대적인 수행에 대해서는 평균으로의 회귀만 작용하고 그 결과 능력이 낮은 사람들의 과대 추정이 줄어들어야 한다.

두 번째로 Burson 등(2006)의 설명은 더욱 간명하다. 이들에 따르면 참가자들의 수행 추정은 과제의 난이도에 따라 달라진다. 즉 과제가 쉬운 경우에는 모두가 자신이 남들보다 더 잘 했으리라고 예상하므로 실제로 수행이 더 높은 사람들(능력이 뛰어난 사람들)은 예측의 정확성이 높고, 과제가 어려운 경우에는 모두가 자신이 남들보다 못 했을 것으로 예상하기 때문에 수행이 낮게 나온 사람들이 결과적으로 예측의 정확성이 더 높게 된다는 것이다. Burson 등(2006)은 이 가설을 검증하기 위해서 과제의 난이도를 조작한 뒤에 능력의 수준이 다른 사람들의 추정의 정확성을 관찰하였다. 실험의 결과 이들의 예상대로 어려운 과제에서는 능력이 높은 사람들이 오히려 더 낮은 정확성을 나타내었다. Burson 등의 해석에 의하면 Kruger와 Dunning(1999)의 실험에서는 비교적 쉬운 과제를 썼기 때문에 능력이 낮은 사람들이 정확성이 낮았던 것이 된다. 하지만 Burson 등(2006)의 연구에서는 사용한 과제들이 대부분 참가자들에게 생소한(노벨상 수상자에 대한 질문들, 시카고 대학교에 관한 퀴즈 등) 것들이었다는 문제가 있다. 아울러 Krueger와 Mueller(2002)의 경우와 마찬가지로 종속 변인은 일관적으로 자신의 상대적 수행(백분

위)이었기 때문에 절대적 수행(점수)에 있어서의 Dunning-Kruger 효과에 대해서는 설명을 하지 못하는 한계를 지닌다.

본 연구에서는 Dunning-Kruger 효과가 한국의 대학교에서 수업의 시험이라는 과제에 대해 나타나는지를 살펴보고, 선행 연구에서 자세히 살펴 보지 않은 몇 가지 이론적인 문제들을 탐구하였다. 구체적으로 이 연구의 목적은 다음과 같다.

첫째, Kruger와 Dunning(1999)의 연구 결과를 한국 대학교라는 문화적인 맥락과 시험 성적이 라고 하는 현실 속의 과제에 대해 반복 검증한다(연구 문제 1). 본 연구를 수행하기 전에 조사한 바에 따르면 Kruger와 Dunning의 연구 결과는 한국에서 검증된 바가 드물다(의학 교육 장면에서 수행된 허정원, 한재진, 임현정, 2008과 한미현, 박석건, 2009 연구 참조). Krueger와 Mueller(2002)의 주장처럼 Dunning-Kruger 효과가 자기 고양 욕구에 일부 기인하는 것이라면, 자기 고양보다는 자신의 능력을 낮춰보는 성향을 보이는 동양권에서는(Heine, Lehman, Markus, & Kitayama, 1999) 능력이 낮은 사람들이 자신의 수행을 과대 추정하는 현상이 훨씬 약하게 나타날 것이다. 하지만 Kruger와 Dunning(1999)의 주장대로 이 현상이 동기적 원인이 아닌 메타 인지에서의 차이 때문에 나타나는 것이라면 한국의 맥락에서도 이 현상이 발견될 것이다.

시험 성적을 추정하는 것은 여러 가지 측면에서 Dunning-Kruger 효과의 보수적인 검증이 된다. 먼저 수업 내용의 시험은 유머 감각(Kruger & Dunning, 1999)이나 카드 놀이인 브릿지(Simons, 2013) 등에 비해 잘 정의되고 객관적으로 수행의 평가가 가능한 과제이며 생태학적 타당성이 높고(Ehrlinger 등, 2008) 당연히 현실적으로 매우 중요하다. 또한 과제 자체의 성격상 난이도가 어느 정도 이상 높으므로, Burson 등(2006)의 주장대로 어려운 과제에서는 능력이 낮은 사람들의 추정 정확성이 높아지는지 검증할 수 있다. 그리고 시험을 보는 것은 대학 생활을 거치면서 여러 번 반복되는 것이므로, 대부분의

학생들이 자신의 학업 능력에 어느 정도 이해를 가지고 있을 것으로 예상할 수 있다. 이런 특성들을 고려하여 Dunning 등(Dunning 등, 2003; Ehrlinger 등, 2008)은 시험의 맥락에서 Dunning-Kruger 효과를 검증하였다. Hacker, Bol, Horgan과 Rakow(2000)은 학생들이 반복된 시험 사이사이에 자신의 수행에 대한 피드백을 받고 문제를 분석할 기회가 주어지는 경우에도 능력이 부족한 학생들의 수행의 과대 추정 현상은 쉽게 사라지지 않음을 보였다. 아울러 Dunning-Kruger 효과에 대한 대부분의 연구에서는 상대적인 수행의 추정에 보다 초점을 맞추고 있는 것과 달리, 이 연구에서는 절대적인 수행의 추정을 종속변인으로 하였다. 그렇게 함으로써 Dunning-Kruger 효과 중에서 타인의 수행의 추정과 관련한 부분을 통제하고 순수하게 자신의 절대적 수행의 추정에 있어서의 오류만을 관찰할 수 있었다.

둘째, 스스로 시험을 준비하면서 들인 노력의 정도가 자신의 과제 수행의 추정에 영향을 미치는지를 살펴본다(연구 문제 2). 노력의 효과는 선행 연구에서 그다지 다루어지지 않은 측면인데, 효과는 두 가지로 생각해 볼 수 있다. 한편으로 시험 공부를 열심히 하는 것은 해당 분야에서의 지식을 증가시킬 수 있고 그럼으로써 자신의 능력에 대한 이해도 깊어질 수 있기 때문에, 노력의 증가는 과대 추정의 감소로 이어질 수 있을 것이다. 하지만 다른 한편 노력을 많이 한 사람은 자신의 노력에 대한 보상 심리 때문에 자신의 수행에 대해 더 긍정적인 결과를 기대할 수도 있을 것이며, 이는 수행의 과대 추정의 증가를 초래할 수 있다. 따라서 이 두 가지 효과들이 모두 존재한다면 서로 반대 방향으로 작용하면서 상충할 것인데, 이 연구에서는 그 가능성을 검토하였다. Hacker 등(2000)은 시험 공부에 더 많은 시간을 투자했을수록 자신의 예상 점수를 높게 본다는 결과를 제시하였지만, 노력이 추정의 정확성에 미치는 영향을 보지는 않았다.

셋째, 과제의 지각된 난이도와 관련한 효과를 살펴본다(연구 문제 3). 위에서 살펴 본 Burson

등(2006)이 맞다면 Dunning-Kruger 효과는 Kruger와 Dunning(1999)의 주장처럼 능력이 뛰어난 사람들이 자신에 대해 더 잘 알아서가 아니라 단순히 모든 사람들이 과제의 지각된 난이도를 기초로 해서 자신의 수행을 추정하기 때문에 일어난 것일 뿐이다. 즉 쉬운 과제에 대해서는 모든 사람들이 자신의 수행을 낙관적으로 추정하므로 실제 능력이 뛰어난 사람들이 더 정확해지게 되고, 어려운 과제에 대해서는 모두가 자신의 수행을 비관적으로 추정하므로 능력이 낮은 사람들이 더 정확해지게 된다. 본 연구에서는 Burson 등과는 다른 방법으로 과제 자체의 난이도를 조작하는 대신 개인적으로 지각한 난이도를 측정하여 그것이 실제 수행과 과대 추정 정도 사이의 관계를 조절하는지를 살펴 보았다. Burson 등이 맞다면 과제를 쉽게 느낄수록 잘한 학생들이 정확하고, 어렵게 느낄수록 못한 학생들이 정확해질 것이다.

넷째, 얼마나 추상적이거나 구체적인 수준에서 추정을 하는지가 오류의 크기에 영향을 미치는지를 살펴 본다(연구 문제 4). Dunning 등(2003)은 사람들이 자신의 수행을 잘 추정하지 못하는 이유가 구체적인 경험을 바탕으로 해서 추정을 하는 상향 처리(bottom-up processing)를 하기보다는 자신에 대해 가지고 있는 기존 신념을 유지한 채로 그것을 수행 예측에 적용하는 하향 처리(top-down processing)를 하기 때문이라고 하였다. Ehrlinger와 Dunning(2003)은 자기 자신에 대한 기존의 평가가 수행의 추정에 영향을 줄 수 있음을 보였다. 예를 들어 한 연구에서는 다양한 영역에 있어서 참가자들이 생각하는 자신의 능력을 평가하게 했다. 그리고 나서 논리에 관한 문제를 푼 뒤에 자신의 수행을 예측하게 하였는데, 이 문제들은 실험 조건에 따라 ‘추상적 추리 능력 시험’ 또는 ‘컴퓨터 프로그래밍 능력 시험’이라고 묘사되었다. 실험 결과 똑같은 문제들을 ‘추상적 추리 능력 시험’이라고 이름 붙였을 때는 ‘컴퓨터 프로그래밍 능력 시험’이라고 했을 때에 비해 그 수행에 대한 추정이 자

신의 추상적 추리 능력에 대한 평가의 영향을 더 많이 받고, 그 반대도 마찬가지임이 밝혀졌다. 즉 수행 자체의 경험 뿐 아니라 수행과 직접 무관한 자신에 대한 평가가 수행의 추정에 영향을 미치는 것이다. 본 연구에서는 수행의 추정에 관한 질문에서 추상성과 구체성을 조작하여 그 효과를 살펴보았다. 해석 수준 이론(Construal Level Theory: Trope & Liberman, 2010)에 의하면 추상적인 수준에서의 해석은 국지적(局部的)이고 맥락화된 정보 처리보다는 보다 전역적(全域的)이고 일반화된 정보 처리로 이어지게 된다. 따라서 추상적인 해석을 유도했을 때는 구체적인 해석을 유도했을 때보다 하향 처리가 더 많이 이루어져서, 시험에서의 구체적 경험을 덜 고려하는 결과 수행의 과대 추정의 경향이 더 증가할 수 있을 것이다. 본 연구에서는 참가자들을 두 조건에 무선 할당하여 자신의 시험 점수를 예측하게 하거나(추상적 추정 조건) 객관식과 주관식 문제를 각각 몇 개나 맞추었는지를 추정하게 하였다(구체적 추정 조건). 문제들을 몇 개 맞추었는지 생각하게 하는 것은 자신의 능력에 대한 전반적인 느낌보다는 개별 문제들에서의 수행에 관한 구체적인 고려를 유도함으로써, 보다 정확한 추정을 하게 할 것으로 예상하였다.

참가자들은 학부 과목의 기말 시험을 본 직후에 자신의 수행을 추정하였는데, 할당된 조건에 따라 구체적 또는 추상적인 수준에서 추정을 하도록 안내되었다. 이 점수는 뒤에 참가자들의 실제 점수와 비교되었다. 참가자들은 또한 자신이 느끼는 시험의 난이도와 자신이 시험을 준비하면서 들인 노력 등에 대해서도 보고하였다.

방 법

연구 대상 및 절차

4년제 종합대학의 심리학 전공과목을 수강하

고 있는 학부생들이 이 연구에 참가하였다. 이 과목은 해당 학과의 2학년 2학기의 전공 선택과목으로 지정된 과목이다. 자료 수집은 기말 고사 시험 직후에 이루어졌다. 참가자들은 이 수업을 듣고 있던 57명(여자 36명) 중에서 자유의사에 의해 설문에 응한 학생들이다. 질문들 중에 하나라도 유효한 답을 하지 않은 학생들은 분석에서 제외하였으며, 그 결과 52명의 자료가 분석에 사용되었다. 답변자들의 성별이나 연령 등의 정보는 수집하지 않았다. 참가자들은 무선적으로 구체적 추정 조건과 추상적 추정 조건에 할당되었으며 시험지의 끝 부분에 있는 질문에 답을 하였다.

측정 도구

참가자들은 다음의 순서로 제시된 질문에 답을 하였다. 첫 번째 질문에서 참가자들은 자신의 수행을 추정하였는데, 참가자들이 할당된 조건에 따라 수행 추정 질문이 달랐다. 구체적 추정 조건의 참가자들은 자신이 맞추었을 것으로 예상하는 문제의 수를 추정하였다(“객관식(30문제)와 주관식(다섯 문제) 각각에 있어서 몇 개나 맞춘 것 같나요?”). 추상적 추정 조건의 참가자들은 시험 점수에 관한 직접적인 질문을 받았다(“기말 고사 점수는 100점 만점에 몇 점 정도 받을 것 같나요?”). 이어서 참가자들은 평균적인 수행 추정의 질문을 받았다(구체적 추정 조건: “다른 학생들은 평균적으로 각각 몇 개 정도 맞출 것 같나요?”; 추상적 추정 조건: “반 평균은 몇 점 정도일 것 같나요?”). 이 문항에 대한 답변은 본 논문의 결과 분석에 포함되지 않았다.

그 다음으로 참가자들은 시험의 난이도에 대한 질문(“시험이 얼마나 어려웠나요?”)에 대해 5점 척도(1 = “아주 쉬웠다”; 5 = “아주 어려웠다” 등)에 평정을 했다. 마지막 질문은 시험 공부에 쏟은 개인적인 노력에 관한 질문(“시험 공부를 얼마나 열심히 했나요?”)이었으며 척도는 1 = “전혀 하지 않았다”부터 5 = “아주 열심히

했다”까지 5점 척도였다.

결 과

점수 계산과 집단 분류

추정치 범위로 제시한 경우(예: “4~5문제”)에는 그 범위의 평균값을 추정치로 간주했고 “모르겠다” 등의 유효하지 않은 답변을 한 경우에는 설문에 응하지 않은 것으로 간주했다. 참가자들의 실제 점수와 추정 점수는 객관식 문제에는 한 개당 2점을, 주관식 문제에는 한 개당 6점을 곱하여 합하였으며, 이 점수에 10/9를 곱함으로써 100점 만점이 되도록 하였다. 추상적 추정 조건에서는 참가자들이 제시한 추정 점수를 그대로 사용하였다.

Kruger와 Dunning(1999)의 방법을 따라 참가자들의 실제 점수를 서열화시켜서 상위부터 25%씩 묶어서 1에서 4분위의 네 집단을 만들었다. 표 1은 각 사분위에 속한 참가자들의 추정 점수와 실제 점수, 그리고 그 차이를 나타내고 있다. 참가자들의 기말 고사의 실제 점수의 평균은 71.8점으로서 난이도가 어느 정도 높았음을 알 수 있다. 또한 1사분위와 4사분위에 해당하는 집단들의 평균 점수의 차이가 42.0점으로 나타나 점수의 변산성이 상당히 높았다. 지각된 난이도의 평균은 3.64점으로 척도의 중간 점수인

3점보다 유의미하게 높았다, $p < .001$. 종합하면 객관적으로나 주관적으로나 대체로 어려운 시험이었음을 알 수 있으며, 이는 쉬운 과제에서는 능력이 높은 사람들이, 어려운 과제에서는 능력이 낮은 사람들이 더 정확할 것이라는 Burson 등(2006)의 주장에 대한 보수적인 검증의 기회를 제공할 수 있을 것이다(유사한 논의로 Ehrlinger 등, 2008을 참조).

연구 문제 1: Dunning-Kruger 효과의 검증

참가자들은 대체로 자신의 점수를 어느 정도 잘 추정하였다. 참가자들이 받은 실제 점수와 그들이 스스로 추정한 점수는 정적 상관을 나타내어 $r = .65$, $p < .001$ 이었다. 앞서 Dunning 등(2003)은 한 연구에서의 실제 점수와 추정 점수의 상관이 $r = .60$ 정도였던 것으로 보고하고 있어, 본 연구에서도 그에 비교할 수 있을 정도로 추정이 정확했음을 알 수 있다. 하지만 이 연구에서의 주된 관심은 그 추정치가 얼마나 부정확한가, 부정확할 경우 그 방향은 어떠한가, 참가자들의 실제 점수가 부정확성과 관련이 있는가이다. 이것을 알아보기 위해 점수를 종속변인으로, 점수가 실제 점수인지 추정 점수인지를 참가자내 요인으로, 각 참가자가 소속된 사분위(1, 2, 3 또는 4)를 참가자간 요인으로 하여 혼합 요인 변량분석을 실시하였다.

먼저 사분위의 주효과가 나타났다, $F(3, 48) = 36.52$, $p < .001$, $\eta^2 = .70$. 또한 실제 점수와 추정 점수의 차이도 유의미하여, $F(1, 48) = 6.66$, $p = .01$, $\eta^2 = .12$, 참가자들은 평균적으로 3.5점 정도 실제보다 높은 점수를 예상하였다. 또한 이런 경향성은 가설에 일치하는 방향으로 사분위에 따라 큰 차이를 보였는데, 참가자내 요인과 사분위의 상호작용이 유의미하였다, $F(3, 48) = 13.49$, $p < .001$, $\eta^2 = .46$. 표 1과 그림 1에 나타난 바와 같이 과대 추정의 주효과는 모두 실제 점수가 가장 낮은 1사분위의 참가자들이 이끌어낸 것이었다. 이들의 과대 추정의 크기는 17.5점이나 되었으며 이 차이는 유의미하였다,

표 1. 각 사분위에서의 추정 점수와 실제 점수, 그 차이값들

추정 점수	실제 점수	차이 (비율%)
66.2 (16.6)	48.7 (11.4)	17.5** (35.9%)
72.3 (9.1)	67.7 (4.0)	4.6 (6.8%)
76.7 (9.0)	79.9 (3.1)	-3.2 (-4.0%)
86.1 (8.6)	90.7 (3.8)	-4.6† (-5.1%)
75.3 (13.2)	71.8 (17.0)	3.5* (4.9%)

주. 괄호 안은 표준편차를 나타냄.

† $p < .10$, * $p < .05$, ** $p < .01$.

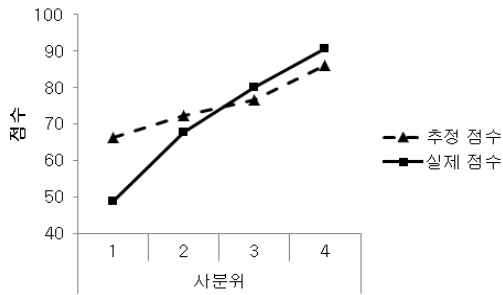


그림 1. 사분위의 함수로 나타낸 추정 점수와 실제 점수

$F(1, 12) = 30.02, p < .001, \eta^2 = .71$. 2사분위의 참가자들도 통계적으로 유의미하지는 않았지만 과대 추정을 하는 경향을 나타내었다. 그에 반해 3, 4사분위의 참가자들은 점수를 과소 추정하는 경향을 보였으며, 4사분위에서의 과소 추정 정도는 $F(1, 12) = 3.63, p = .08, \eta^2 = .23$ 로 유의도에 근접하였다. 이런 일련의 패턴은 Kruger와 Dunning(1999)의 결과와 일치하는 것이다. Tukey의 방법에 의해 사분위간의 추정 오차의 차이에 관한 사후 검정을 한 결과, 1사분위에서의 추정의 오차는 다른 사분위에서의 오차와 유의미한 차이가 있었으며($ps < .05$), 2사분위와 4사분위의 추정의 오차의 차이는 유의도에 근접하였다, $p = .10$.

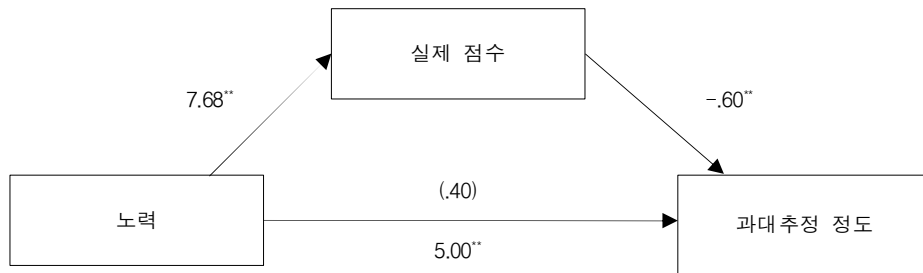
Kruger와 Dunning(1999)을 비롯한 선행 연구들에서는 위와 같이 참가자들을 석차에 따라 네 집단으로 나누어 각 집단을 비교했지만, 과대 추정의 경향은 실제 점수를 연속 변인으로 취급해서도 살펴볼 수 있다. 이를 위해 과대 추정의 정도를 (추정 점수 - 실제 점수)로 정의한 뒤에 이것과 실제 점수의 상관을 검토하였다. 그 결과 $r = -.64, p < .001$ 로 나타나 점수가 낮을수록 과대 추정의 정도가 더 증가한다는 예측을 뒷받침하였다.

연구 문제 2: 과제의 수행 준비에 들인 노력은 능력의 추정에서의 편향에 어떤 영향을 미치는가?

두 번째 연구 문제, 즉 노력이 능력의 추정에

어떤 방향으로 영향을 미치는지를 알아보기 위해, 참가자들이 시험 준비에 들인 노력을 측정하여 이것이 과대추정 정도에 미치는 효과를 살펴 보았다. 먼저 사분위를 독립변인으로 하여 일원변량분석을 한 결과 각 사분위에서의 노력의 차이는 유의미수준에 근접한 효과를 나타내었다, $F(3, 48) = 2.80, p = .05$. 1사분위부터 4사분위까지 노력의 평균이 차례로 3.08, 3.62, 3.77, 4.00으로 나타나, 자신이 쏟았다고 지각하는 노력은 실제 점수에 따라 대략적으로 선형적으로 증가함을 알 수 있다. 실제 점수와 노력간의 상관관은 $r = .40, p = .003$ 으로 나타나서 역시 같은 해석이 가능했다.

노력이 클수록 자신의 수행을 과대추정할 것이라는 기대와 달리 참가자가 시험에 들인 노력은 과대추정 정도와 직접적인 상관관이 없었다, $r = .03, p = .85$. 하지만 앞서 본 것처럼 노력은 실제 점수와 양의 상관관계를 가지고, 또한 점수가 높을수록 과대추정 정도가 낮아지는 것을 고려할 때 실제 점수는 노력이 수행의 과대추정 정도에 미치는 효과에 억제변인(suppressor variable)으로 작용할 가능성을 상정할 수 있다. 이 가능성을 검증해보기 위해 Hayes(2013)가 개발한 SPSS 매크로인 PROCESS를 이용하였다. PROCESS는 bootstrapping 방법으로 다양한 매개 분석을 할 수 있는 기능을 가지고 있다. 이 분석에서는 노력을 독립변인으로, 과대추정 정도를 종속변인으로, 실제 점수를 매개변인(억제변인)으로 하여 회귀분석을 수행하였으며, bootstrapping sampling은 1000번으로 하였다. 이 결과는 그림 2와 같았다. 노력과 실제 점수를 모두 투입해서 실제 점수가 과대추정 정도에 미치는 효과를 통제했더니 노력이 과대추정 정도에 미치는 정적 효과가 나타났다. 노력이 실제 점수를 통해 과대추정 정도에 미치는 간접 효과의 비표준화된 계수는 -4.60 이며 이 계수의 95% 신뢰구간은 -7.85 에서 -1.91 까지로 0을 포함하지 않으므로 유의미했음을 알 수 있다. 정리하면 노력은 실제 점수의 증가로 이어지므



주. 숫자들은 비표준화된 회귀계수들임. 괄호 안은 실제 점수가 투입되기 전에 노력의 과대추정 정도에 대한 효과를 나타낸 것임. ** $p < .01$.

그림 2. 노력과 실제 점수, 과대추정 정도 사이의 매개 분석의 결과

로 과대추정 가능성을 억제하지만, 다른 한편으로 노력을 많이 기울였을수록 수행을 과대추정하는 정도도 증가하는 패턴이 나타났다.

연구 문제 3: 과제의 지각된 난이도는 능력의 추정에서의 편향에 어떤 영향을 미치는가?

이 연구의 세 번째 관심사는 각 참가자들이 지각하는 시험의 전체 난이도가 능력의 추정에 미치는 영향을 살펴보는 것이다. 먼저 지각된 난이도와 실제 점수와의 상관은 $r = .02, p = .90$ 로 유의미하지 않았다. 즉 낮은 점수의 사람들이 문제를 더 어렵게 느끼지는 않은 것이다. 능력이 낮은 사람들이 자신의 능력을 정확히 인지하고 있다면 이 둘의 관계는 양의 상관을 나타내었을 것이므로 이러한 상관의 부재는 Kruger와 Dunning(1999)의 주장을 간접적으로 뒷받침한다고 볼 수 있다.¹⁾ 한편 지각된 어려움은 예상 점수와 $r = -.21, p = .13$ 의 관계를 나타내어 시험을 어렵게 느꼈을수록 예상 점수를 낮게 추정하는 경향성을 드러내었다.

과제의 지각된 난이도가 능력과 추정의 정확성의 관계를 조절할 것이라는 것은 Burson 등(2006)의 연구의 결과이다. 앞서 본 바와 같이 본 연구에서 사용한 시험은 평균적으로 중간보

다 약간 어려운 과제로 판단되었으나, 그럼에도 수행이 좋은 참가자들이 자신의 수행의 추정에 있어서 더 정확했다. 이 결과는 어려운 과제에서는 능력이 낮은 사람들이 오히려 더 정확할 것이라는 Burson 등의 주장과 일치하지 않는다.

이 문제를 보다 상세히 살펴 보기 위해 지각된 난이도와 실제 점수가 추정의 정확성에 상호작용 효과를 일으키는지를 살펴보았다. Burson 등(2006)은 과제의 난이도를 조작해서 그에 따라 능력과 추정의 정확성의 관계가 달라짐을 보았지만 이 효과는 개인 수준에서도 살펴볼 수 있다. 즉 Burson 등의 말대로 어려운 과제에서는 모두 자신의 수행을 부정적으로 추정하기 때문에 능력이 낮은 사람들이 더 정확해지고 쉬운 과제에서는 반대의 이유로 능력이 높은 사람들이 더 정확해지는 것이라면, 지각된 난이도와 실제 점수는 정확성에 음의 방향의(또는 추정 오류에 양의 방향의) 상호작용을 일으켜야 한다. 분석을 하기에 앞서 각 회귀에서 지각된 난이도와 실제 점수를 표준화하고 나서 둘의 곱을 계산하여 상호작용 항을 계산했다. 그리고 이 세 변인들을 독립 변인으로 하고 추정의 오류(실제 점수와 추정 점수의 차이의 절대값)를 종속 변인으로 하여 다중 회귀 분석을 실시하였다. 이 분석의 결과 실제 점수의 부적 효과만이 나타났을 뿐($\beta = -0.36, p = .01$), 지각된 난이도($\beta = -0.10, p = .43$)나 두 변인의 상호작용 효과($\beta =$

1) 지각된 난이도에 있어서의 사분위간 차이를 보기 위한 변량분석의 결과 유의미한 사분위의 효과가 나타나지 않았다, $F(3, 48) = 2.67, ns$.

0.10, $p = .48$)는 관찰되지 않았다. 즉 지각된 난이도에 따라 능력이 높고 낮은 사람들의 정확성에 차이가 나타나지 않았다. 비록 Burson 등(2006)에서와는 달리 여기서는 과제의 난이도를 조작하는 대신 각 참가자들의 지각된 난이도를 측정하였지만, 이 결과는 Dunning-Kruger 효과가 과제의 지각된 난이도의 인위물(artifact)에 불과하다는 Burson 등(2006)의 주장을 적어도 뒷받침하지는 않는다.

연구 문제 4: 추정의 구체성은 능력이 낮은 사람들의 과대 추정 정도를 감소시키는가?

네 번째로, 추정의 구체성 정도가 Dunning-Kruger 효과를 조절할 것인지의 여부를 알아보았다. 즉 보다 추상적인 수준에서 추정을 하면 능력이 낮은 사람이 자신의 수행을 과대 추정하는 정도가 더 강할 것이란 예상이 가능하다. 그런데 참가자들을 두 실험 조건과 네 사분위로 구분해서 분석을 할 경우 각 셀의 크기가 너무 작은 문제가 있어, 실제 점수를 연속 변인으로 취급해서 두 실험 조건에서의 실제 점수와 과대 추정 정도의 상관을 비교하였다. 상관을 보기에 앞서 두 조건에서 예상 점수의 차이가 있는지를 확인해 보았는데 구체적 추정 조건($M = 73.87$)에서보다 추상적 추정 조건($M = 76.75$)에서의 추정 점수가 약간 높긴 했지만 유의미한 차이는 나타나지 않았다, $F(1, 50) = .61, p = .44$.

구체적 추정 조건에서 두 변인의 상관은 $r = -.58, p = .002$ 이었고, 추상적 추정 조건에서는 $r = -.74, p < .001$ 로 그 값이 약간 더 컸다. 하지만 r -to- z 변환 뒤에 비교를 하면 $z = 1.00, p = .32$ 로 두 상관은 유의미한 차이를 보이지 않았다. 즉 이 조작은 기대했던 조절 효과를 유발하지 않았다. 이 조절 효과의 부재를 재확인하기 위해서 두 번째 방법으로 실제 점수와 실험 조건, 그리고 그 상호작용 효과가 과대 추정 정도를 예측하는 회귀 분석을 시행하였다. 이 분석의 결과 실제 점수의 주효과($\beta = -.64, p < .001$)만이 관찰되었을 뿐 실험 조건의 주효과(β

$= .04, p < .76$)나 상호작용 효과($\beta = -.01, p = .95$) 모두 관찰되지 않았다. 이 두 분석의 결과를 종합해 볼 때, 적어도 이 실험에서 사용한 추상성/구체성의 조작으로는 과대 추정 정도에 영향을 미치지 못했음을 알 수 있다.²⁾

논 의

이 연구에서는 참가자들이 매우 익숙한 대학교 수업의 시험이라는 맥락에서 명백한 Dunning-Kruger 효과를 보였다. 이 연구에서는 상대적 수행(백분위)의 추정에 있어서 이 효과를 관찰한 대부분의 연구와 달리 절대적 수행(자신의 점수)의 추정을 종속변인으로 해서 효과를 발견하였다. 구체적으로 가장 능력이 낮은 참가자들이 실제 자신이 받은 점수보다 유의미하게 더 많이 얻을 수 있을 것으로 예상하였고, 상위권의 참가자들은 반대로 실제보다 과소 추정을 하는 경향을 보였다.

이러한 결과는 Krueger와 Mueller(2002) 그리고 Burson 등(2006)의 연구에서 제시한 가설로는 설명하기 어렵다. 즉 그들의 가설은 모두 ‘타인과의 상대적 비교’에 있어서의 과대 추정 효과만을 설명하는 것이므로 절대적인 점수 추정에서 왜 여전히 능력이 낮은 사람들의 과대 추정이 두드러지는지는 잘 설명하지 못한다. Krueger와 Mueller(2002)의 설명에 따르면 Dunning-Kruger 효과는 평균으로의 회귀와 평균 이상 효과가 합하여 나타난 현상이다. 그렇다면 절대적 수행을 추정하게 한 본 연구에서는 평균 이상 효과가

2) 이와 유사한 논리로, 주관식 문제에 대한 하향 처리가 더 강하기 때문에 두 문제 유형에 있어서 능력이 낮은 참가자들의 과대 추정이 차이가 있을 가능성을 생각해 볼 수 있다. 두 유형의 문제의 실제 점수들을 과대 추정 정도와 상관을 계산한 결과 객관식 문제에 대해서는 $r = -.53, p = .006$, 주관식 문제에 대해서는 $r = -.59, p = .001$ 이었으며 이 두 상관계수는 유의미한 차이가 없었다, $z = 0.34, p = .73$.

나타날 수 없으므로 능력이 낮은 사람들의 과대 추정과 능력이 높은 사람들의 과소 추정이 서로 균형을 보여야 하는데, 실제로는 능력이 낮은 사람들의 과대 추정이 더 크게 나타났다. 아울러 지각된 어려움과 실제 점수가 추정 오류에 대해 유의미한 상호작용 효과를 나타내지 않는 것도 Burson 등(2006)의 설명과는 맞지 않는다. 이런 결과들을 봤을 때 본 연구 결과는 Kruger와 Dunning(1999)의 메타 인지 설명에 더 잘 부합하는 것이라고 볼 수 있다. 다만 흥미로운 점은 능력이 뛰어난 사람들에 대한 결과인데, 점수가 높은 사람들이 자신의 수행을 과소 추정하는 경향을 보였다. 앞서 본 것처럼 Kruger와 Dunning(1999)은 능력이 뛰어난 사람들의 과소 추정은 타인과의 상대적 비교에서의 오류라는 견해를 제시하고 있다. 이에 따르면 이와 같은 과소 추정은 자신의 상대적인 수행(즉 전체에서의 석차)의 추정에서만 나타나야 하며 본 연구에서와 같이 절대적인 수행을 추정할 때는 나타날 수 없으므로 본 연구에서의 점수가 높은 사람들에 대한 패턴은 설명하기가 쉽지 않다. 한국인들의 겸손한 성향(Heine 등, 1999)에 귀인을 하는 것이 하나의 가능한 설명이 되겠지만, Ehrlinger 등(2008)이 미국에서 실행한 연구들 중 일부에서도 능력이 뛰어난 사람들이 절대적 점수를 과소 추정할 바 있기 때문에 그 설명을 적용하기가 쉽지 않다. Ehrlinger 등(2008)도 이 문제에 대해서는 명확한 설명을 제시하지 못하고 있는 바, 이 문제는 앞으로 연구가 더 필요하다.

학업에 있어서 Dunning-Kruger 효과가 관찰된 것은 학생들의 학습 효과를 증진시키는 측면에서 우려스러운 결과이다. 대학교 학부생들은 매 학기에 수 차례 시험을 보며 시험에 대해 직접적인 피드백을 받을 기회도 많다. 따라서 한두 학기가 지나면 - 본 연구의 자료 수집에 사용된 수업은 2학년 2학기 수업이었다 - 자신의 학업 능력에 대해 어느 정도의 이해를 갖추고 있을 것으로 예상할 수 있지만 결과는 그렇지 못하였다. 같은 수업에서 이미 중간 고사를 본 뒤에도

추정이 나아지지 않은 것을 보면 아마도 학업 능력이 낮은 학생들은 자신의 정서 조절을 위해서 자신의 시험 성적에 대한 피드백도 잘 받으려 하지 않으려 한다는 추측을 해 볼 수는 있다. 그 결과 학업 능력이 높은 학생들은 계속 높은 수준을 유지하고, 떨어지는 학생들은 낮은 수행 수준에서 좀처럼 벗어날 수 없게 될 수 있다. 따라서 학업 능력이 떨어지는 학생들이 자신의 능력에 대해 메타 인지를 더 잘 해서 보다 적절하게 노력을 할 수 있도록 하는 것은 매우 강한 현실적인 함의를 지닌다.

둘째, 참가자 스스로 지각한 본인의 노력은 수행의 과대 추정 정도에 대해 서로 엇갈리는 직접 효과와 간접 효과를 작용시켰다. 이 결과는 노력을 하면 할수록 실제 성적이 좋아지는 한편 자신의 능력에 대한 이해도 높아져서 수행의 추정이 더 정확해지지만(Dunning-Kruger 효과), 동시에 자신이 들인 노력만큼 수행이 좋아질 것으로 예상하기도 하므로 수행을 과대 추정하는 결과로도 이어지는 것으로 해석할 수 있다. 결과적으로 이 두 효과는 서로 반대 방향으로 작용하여(즉 실제 수행을 통한 효과가 노력의 직접 효과에 대한 억제 효과의 역할을 해서) 표면적으로는 노력이 과대 추정 정도와 별다른 상관관계를 나타내지 않는다. 특히 후자인 노력 --> 수행 과다 추정의 효과는 선행 연구에서 그다지 보고되지 않은 효과로 보이는데, 이 효과의 원인이 동기적인 데(예: 자신의 투자만큼 보상을 얻고 싶은 마음) 있는지 메타 인지적인 데(예: 자신이 들인 인지적인 노력의 양만큼 지식이 증가했을 것이라는 생각) 있는지는 앞으로 더 연구해 볼 흥미로운 주제라고 생각된다. 현실적인 측면에서 이 결과는 노력을 통해 자신의 능력을 이해하는 것이 간단한 문제가 아님을 보여 준다.

셋째, 난이도 관련 결과는 Dunning-Kruger 효과에 대한 Burson 등(2006)의 설명을 뒷받침해주지 않았다. 즉 Burson 등의 주장에 따르면 지각된 난이도가 높을수록 실제 점수와 수행의 과

다 추정 사이의 음의 효과가 더 강해져야 하지만, 그러한 상호작용은 나타나지 않았다. 본 연구에서는 Burson 등(2006)과는 달리 과제의 난이도를 조작하기보다는 각 참가자들이 지각한 난이도를 측정했다는 것을 분명히 할 필요가 있다. 그런데 그들의 설명에서는 사람들이 문제의 지각된 난이도를 자신의 상대적인 수행의 척도로서 사용하기 때문에 과제가 쉬우면 능력 높은 사람들의 추정이, 과제가 어려우면 능력 낮은 사람들의 추정이 정확해질 거라고 한 것이다. 따라서 주관적인 난이도를 사용해서도 유사한 결과가 나오길 기대할 수 있다. 하지만 선행 연구와 다른 방법을 썼으므로 이 결과는 유의해서 검토할 필요가 있다.

넷째, 자신의 수행의 추정을 구체적으로 또는 추상적으로 묻는 조작은 추정 패턴에 차이를 가져오지 않았다. Dunning 등(2003)에 의하면 사람들이 자신의 실제 수행 정보를 수행의 추정에 반영하지 못하는 것은 지나친 하향처리 때문인데, 이 조작은 상향처리와 하향처리의 정도에 변화를 주어서 추정의 정확성을 증진시키는 전략을 발견하려는 시도였으나 성공하지 못했다. 구체성/추상성의 정도의 차이가 충분히 크지 않아, 두 종류의 질문이 거의 같아서 그런 것이었을 가능성이 있다. 즉 추상적 추정 조건에서 점수를 추정할 때 자신이 몇 개의 문제를 맞추었는지를 추정하고 그 추정치로부터 예상 점수를 계산했기 때문에 조건간 차이가 나타나지 않았을 것이라 추측을 할 수 있다. 두 조건에서 서로 다른 과정을 거치는 조작을 더 강하게 했다면, 예를 들어 추상적 추정은 보다 전체적인 느낌을 강조한 질문을 하고 구체적 추정에서는 각각의 문제들을 맞혔을지를 따져서 합쳐 보게 하면 더 효과적인 조작이 됐을 수도 있다. 앞으로의 연구에서는 수행의 추정은 그대로 두고 해석 수준 이론(Trope & Liberman, 2010)의 연구에서 사용하는 추상성 조작 방법들을 사용해서 이 가능성을 검증할 수도 있을 것이다. 이 방법들은 크게 과제 자체와 무관하게 사고의 틀(mindset)을

점화하는 방법(예: 시간적/공간적 거리의 조작; 여러 사례들을 묶는 범주 이름 제시하기 또는 주어진 범주의 예를 제시하기 등)이 있고 과제 자체에서 조작을 하는 방법이 있는데, 후자 쪽의 한 예로서 수행을 추정하는 질문에 추상적이거나 구체적인 표현들을 사용하는 것을 고려할 수 있다.

연구의 한계와 후속 연구의 방향

이 연구의 한계들 중에 가장 두드러진 것은 적은 표본 크기이다. 참가자 수가 적었기 때문에 구체성 조작의 효과 검증에 한계가 있을 수밖에 없었다. 설문을 간단히 하기 위해 각 변인들을 측정하는 질문들이 하나씩밖에 없었기 때문에 신뢰도의 측면에서 문제가 있을 수밖에 없다. 그리고 선행 연구들에서 측정했던 상대적 수행의 추정이나 확신 등을 설문에 넣었다면 더 다양한 대안 설명들을 보다 체계적으로 검증해 볼 수 있었을 것이다.

이 연구에서 검토한 네 가지 연구 문제들 중에 난이도에 관한 것과 추상성/구체성에 관한 것에서는 유의미한 효과를 얻지 못했다. 특히 실제 수행 --> 수행의 추정의 관계에 대한 지각된 난이도의 조절 효과를 관찰하지 못한 것은 Burson 등(2006)의 설명에 배치하는 결과로서 이 설명의 타당성에 의문을 주고 다른 대안 설명들에 힘을 더 실어 주는 결과라고 할 수 있다. 물론 영가설을 기각하지 못한 것에 대해 조작의 실패 등의 다른 원인을 생각할 수 있으므로 해석에 각별히 주의가 필요하다. 그럼에도 불구하고, Ehrlinger 등(2008)도 Dunning-Kruger 효과에 관한 자신들의 설명을 뒷받침하는 증거로서 여러 조작들의 유의미하지 않은 효과들을 제시하였음을 고려할 때 본 연구에서의 유의미하지 않은 결과들도 검토를 할 가치가 있다고 생각된다. 일찍이 Greenwald(1975)가 영가설을 받아들이는 결과들도 충분히 학문적 기여의 여지가 있을 수 있음을 역설한 것에 비추어, 본 연구에서 보

고된 유의미하지 않은 결과들도 - 추후에 체계적인 반복 검증이 이루어질 것을 전제로 - 잠정적으로 경험적 증거로서의 가치를 인정할 필요가 있을 것이다.

수행 추정 질문과 시험 답안지가 서로 연결이 되어 있기 때문에 조사의 익명성을 확보할 수 없었던 것도 이 연구의 한계이다. 이 문제는 생태학적 타당성을 위해 실제의 시험에서의 수행을 연구 대상으로 한 것의 불가피한 결과로서, Burson 등(2006)이나 Ehrlinger 등(2008)의 선행 연구에도 해당되는 한계이다. 특히 능력이 낮은 참가자들은 자신을 과대평가하기를 꺼려서 더 정확히, 또는 더 낮게 추정을 했을 수도 있고, 또는 반대로 자신이 현실적으로 얻을 점수를 알면서도 체면 때문에 점수를 적절히 높게 추정했을 가능성도 있다. 연구자가 참가자들의 개인 정보를 확인할 수 없더라도, 수행 추정의 답변이 성적에 영향을 미치지 않는다는 지시문이 있었다면, 익명성을 보장할 수는 없더라도 참가자들로부터 연구자의 시선을 덜 의식하는 답변을 얻을 수 있었을 것이다. 참가자들이 들인 노력의 자기 보고에 있어서도 같은 문제가 발생했을 수 있는데, 이 경우 공부에 투자한 시간 등의 보다 절대적인 측정을 통해 왜곡을 줄일 수 있었을 것이다.

이 연구에서는 살펴보지 않았지만, Dunning-Kruger 효과에 대한 인상 관리나 자기 보호 측면에서의 대안 설명도 앞으로 검토해야 할 것이다. 이 설명에 의하면, 능력이 높은 사람들은 교만한 태도를 취하거나 교만하게 보이기를 꺼려서 자신의 수행을 낮게 예측하고, 능력이 낮은 사람들은 자신의 능력 부족을 스스로, 또는 공적으로 인정하고 싶지 않아서 자신의 예상보다 더 높은 답변을 하는 것이라는 설명이 가능하다. 이 가능성은 메타 인지 설명과 반드시 배치되는 것은 아니며, 메타 인지에 의한 효과와 동시에 일어나는 것일 수도 있다. Ehrlinger 등(2008)은 이런 가능성을 비롯한 동기적 설명들을 배제하기 위해서 금전적인 인센티브나 설명의

무(accountability)의 조작을 한 뒤에도 여전히 Dunning-Kruger 효과가 나타남을 보였다. 하지만 이런 인상 관리의 동기가 실험에서 제시된 인센티브 등의 효과보다 더 강력했기 때문에 이런 조작의 효과가 없었던 것일 가능성도 있다. 후속 연구에서는 답변을 연구자가 확인할 수 없거나 하는 지시문을 통해 인상 관리의 동기를 줄인다거나, 자기 가치 확인(self-affirmation; Sherman & Cohen, 2006) 조작으로 자신에 대한 위협에 덜 예민하게 반응하도록 하는 등의 방법으로 동기적인 설명을 검증해 볼 수 있을 것이다.

Dunning-Kruger 효과에 대한 대부분의 연구에서는 즉각적인 자기 지각에서의 결과를 조명하였지만, 이 효과로 인해 어떤 장기적인 결과가 나타나는지를 살펴볼 필요가 있다. Mattern, Burrus, 그리고 Shaw(2010)의 최근 연구에 따르면, 역설적으로 자신의 수행에 대해 지나치게 낙관적인 기대를 하는 사람들은 오히려 대학교에서 학점이 더 좋고 졸업을 할 가능성도 더 높다고 한다. 하지만 Heine 등(1999)의 주장대로 동양 문화권의 사람들이 자신에 대한 부정적인 평가로부터 오히려 더 동기화가 된다면 한국을 비롯한 동양에서는 반대로(그리고 Dunning 등의 주장대로) 과대 추정의 결과 학업 수행의 장기적인 저하가 일어날 것을 예상할 수 있다. 실제로 이런 Dunning-Kruger 효과의 부작용이 확인된다면, 결론적으로 이 연구에서는 한국의 학부 대학생들을 대상으로 학업의 맥락에서 Dunning-Kruger 효과를 보여 주었다. 이 효과의 메커니즘의 측면에 있어서는 평균 이상 효과나 난이도에 의한 설명 등보다는 메타 인지에 의한 설명에 보다 일치하는 결과가 나타났다. 과제 수행에 들인 노력은 수행의 과다 추정을 증가시키는 직접 효과와 감소시키는 간접 효과를 모두 일으키는 패턴을 나타내었다. 추상성의 조작은 효과에 차이를 유발하지 않았다. 능력이 낮은 사람은 자신이 얼마나 능력이 낮은지 모른다는 얼핏 보면 당연해 보이는 이 현상은, 능력의 차이가 영속

화될 가능성을 함의하고 있기 때문에 결코 사소하게 생각해서는 안 된다. 따라서 앞으로 이론적으로는 이 효과의 다양한 잠재적 설명들에 대해 보다 종합적인 연구가 이루어져야 할 것이며, 현실적으로는 교육에 있어서 능력에 대한 메타 인지를 증진시키는 학습법의 개발을 위해 고민이 있어야 할 것이다.

참고문헌

- 한미현, 박석건. (2009). 객관구조화진료시험에서 평가자 평가와 수험자의 자기평가 간의 차이. *한국의학교육*, 21(3), 279 - 285.
- 허정원, 한재진, 임현정. (2008). 자기주도학습을 도입한 임상실습에서 의과대학생이 설정한 학습목표, 방법 및 의견과 자기평가 점수의 분석. *한국의학교육*, 20(2), 123-132.
- Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, 90, 60-77.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance, *Advances in Experimental Social Psychology*, 44, 247-296.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83-87.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84(1), 5-17.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105, 98-121.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1-20.
- Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92, 160-170.
- Hayes (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis*. New York: Guilford Press.
- Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106, 766-794.
- Krueger, J. (1998). Enhancement bias in the description of self and others. *Personality and Social Psychology Bulletin*, 24, 505-516
- Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The contribution of social-perceptual skills and statistical regression to self-enhancement biases. *Journal of Personality and Social Psychology*, 82, 180-188.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134.
- Mattern, K. D., Burrus, J., & Shaw, E. (2009). When both the skilled and unskilled are unaware: Consequences for academic performance. *Self and Identity*, 9, 129-141.
- Parikh, A., McReelis, K., & Hodges, B. (2001). Student feedback in problem based learning: a survey of 103 final year students across five Ontario medical schools. *Medical Education*, 35,

- 632-6.
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.). *Advances in Experimental Social Psychology* (Vol. 38, pp. 183-242). San Diego, CA: Academic Press.
- Simons, D. J. (2013). Unskilled and optimistic: Overconfident predictions despite calibrated knowledge of relative skill. *Psychonomic Bulletin & Review* (2013), 1-7.
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440-463.
- 1 차원고접수 : 2013. 04. 14.
수정원고접수 : 2013. 07. 14.
최종게재결정 : 2013. 08. 12.

Unawareness of One's Own Task Incompetence: Replication of Kruger and Dunning (1999) in the Academic Domain

Sang Hee Park

Department of Psychology, Chungbuk National University

Kruger and Dunning (1999) demonstrated that those who are less competent in a specific domain are more likely to overestimate their performance. In the current research, this so-called Dunning-Kruger effect was replicated in a real-life academic context with data collection in an undergraduate course. Right after the final exam, students estimated their performance. Comparisons of these estimates with actual scores revealed Dunning-Kruger effect. A pattern emerged in which, effort in exam preparation had both an indirect, decreasing effect *and* a direct, increasing effect on overestimation. An alternative explanation of Dunning-Kruger effect that it is an artifact caused by perceived difficulty was not supported, and whether performance estimation was made at a more specific or a more abstract level did not lead to a significant difference in the degree of overestimation. In sum, the results of this study is more consistent with Kruger and Dunning's (1999) 'metacognition' account than either 'regression to the mean' account (Krueger & Mueller, 2002) or 'task difficulty' account (Burson, Larrick, & Klayman, 2006). Alternative explanations, limitations of the study, and future directions are discussed.

Key words : performance; metacognition; self-evaluation; overconfidence