

자기보고식 괴롭힘 경험률 평가의 편향요인 탐색: 평가조건 변인을 중심으로*

이 동 형[†]

부산대학교 교육학과

국내외의 학교폭력이나 괴롭힘 실태조사에서 괴롭힘 경험률을 파악하기 위해 자기보고식 평가방법이 널리 활용되는데 불구하고, 자기보고에 기초한 가해율 및 피해율이 구체적인 조작화 방식, 측정방법 등에 따라 적지 않은 차이를 보인다는 지적이 제기되면서 괴롭힘 평가에서 편향요인은 주요 쟁점이 되어왔다. 이러한 배경에서 본 연구는 690명의 남녀 중학생을 지시문을 통해 조작된 서로 다른 네 평가조건(정의제시/미제시, 익명/기명실시)에 무선배정한 후, Olweus 괴롭힘 질문지의 가해 및 피해문항을 실시하여 성별에 따른 경험률의 차이를 살펴보고, 정의제시 및 익명실시 여부에 따라, 그리고 질문형태(전반질문/세부질문)에 따라 경험률이 다르게 나타나는지, 두 가지 절단점과 성별을 고려하여 살펴보았다. 교차분석 결과, 전반질문에서는 남학생이 여학생에 비해 가해율 및 피해율이 높았고, 직접적 괴롭힘에 대한 세부질문에서도 남학생의 경험률이 유의하게 높았지만, 관계적 괴롭힘에서는 성차가 나타나지 않았다. 또한 괴롭힘에 대한 정의가 제시되지 않은 경우, '엄격한' 절단점을 사용할 때, 전반적 가해 경험률뿐 아니라 언어적 괴롭힘 가해율 및 피해율이 11%~15%나 더 높게 나타났다. 그러나 익명실시 여부에 따라서는 한 문항을 제외하고는 경험률에서 차이가 없었으며, 솔직히 응답한 정도나 비밀보장의 확신 정도도 두 조건 간에 다르지 않았다. 질문형태에 따라서는 세부질문을 통해 경험률을 측정할 경우 전반질문 대비 68%~148% 높게 나타났으며, 질문형태에 따른 괴롭힘 경험 측정치 간의 상관도 낮거나 중간 정도 수준이었다. 또한 전반질문은 민감도는 높지 않은 반면, 이에 비해 특이도는 높은 것으로 나타났다. 이러한 결과가 학교폭력이나 괴롭힘 실태조사 문항의 개발과 실시 절차의 개선을 위해 갖는 구체적 시사점을 제시하였으며, 본 연구의 한계점과 의의를 논의하였다.

주요어 : 괴롭힘, 자기보고식 평가, 학교폭력, 편향요인

* 이 논문은 2017년도 부산대학교 인문사회연구기금의 지원을 받아 연구되었음.

[†] 교신저자(Corresponding Author) : 이동형, 부산대학교 사범대학 교육학과, 교수, 부산시 금정구 부산대학교로 63번길 2 / Tel : 051-510-2638, E-mail : leepsy@pusan.ac.kr

학교현장에서 학생들 간에 발생하는 폭력 중 상당 부분을 차지하는 것이 괴롭힘(bullying)이다(Batsche, 1997). 괴롭힘에 대한 학술적 연구는 1960년대 스웨덴의 Daniel Olweus가 시작한 종단연구에서 그 기원을 찾을 수 있지만, 국제적으로 연구물이 급증한 것은 괴롭힘의 치명적 결과가 여러 나라의 사례들을 통해 널리 알려지게 된 1990대 이후였다. 국내의 경우도 대략 2000년대 들어서면서 학교폭력이 교육현장의 주요 문제, 더 나아가 심각한 사회병리적 현상으로 인식되면서 괴롭힘이나 학교폭력에 대한 연구가 본격화되었고, 교육부의 학교폭력 실태 전수조사도 2012년에 도입되어 매년 학생들의 학교폭력 경험률(가해율 및 피해율) 파악하는 데 주요 지표로 활용되고 있다.

괴롭힘은 국내 연구자들 사이에서 집단따돌림, 왕따, 집단괴롭힘, 또래괴롭힘 등과 같은 유사한 용어로도 지칭되며, 보다 포괄적이며, 법률적, 관습적인 용어인 학교폭력과는 동일 시되곤 한다(김혜원, 2013; 이규미 등, 2014; 이동형, 2014a). 그러나 괴롭힘은 의도적/주도적 공격행동이며, 지속성/반복성을 특징으로 하고, 피해학생이 자신을 스스로 방어할 수 없는 힘의 불균형 상태에서 공격행동이 발생한다는 면에서 학교폭력 같이 보다 일반적이며 포괄적인 용어와는 구분된다(Olweus, 2010). 괴롭힘과 학교폭력은 개념상 서로 중첩되는 부분이 있지만, 학교폭력에 포함되는 것으로 간주되는 일부 공격행동의 경우 괴롭힘의 핵심 특성을 갖추지 않았다면, 괴롭힘으로 보기는 어렵다. 즉 괴롭힘은 학생들 간에 나타날 수 있는 다양한 형태의 공격행동(peer aggression 혹은 peer victimization)이나 폭력행위 중 의도성, 반복성, 힘의 불균형의 요소를 갖춘 공격행동을 지칭하는 보다 특정적 개념

이라고 할 수 있다(Jia & Mikami, 2018; Volk, Dane, & Marini, 2014).

괴롭힘의 이러한 핵심요소를 고려할 때, 이를 잘 반영한 괴롭힘 가해 및 피해경험을 측정하는 것은 경험률 집계를 위한 실태조사나 괴롭힘과 관련되는 변인들을 밝히는 과학적 연구, 괴롭힘에 대한 개입이나 예방 프로그램의 효과성을 평가하는 연구 등에서 필수적이라고 할 수 있다. 괴롭힘 관련 경험의 정확한 평가를 위해서는 면접, 관찰, 교사보고, 또래보고, 자기보고 등을 활용한 다면적 평가가 가장 이상적이지만(이동형, 2014b; Crothers & Levinson, 2004), 현실적 제약과 편리성, 경제성 등의 이유로, 특히 괴롭힘 경험률을 평가하고자 할 때는 자기보고식 평가방법이 현재 가장 널리 활용되고 있다(Sveinsson & Morris, 2007). 괴롭힘의 자기보고식 평가에서는 질문지(surveys)나 척도(scales)가 활용되는 데, 경험률 파악을 위한 실태조사에서는 경험 여부(빈도)를 표기하는 방식의 질문지가 주로 사용되고 있다. 이러한 질문지의 대표적인 예는 Olweus Bullying Questionnaire(OBQ: Olweus, 2007)로, 이 OBQ는 현재 다양한 언어로 번역되어 미국을 포함한 많은 국가의 각종 괴롭힘 관련 실태조사에서 그 질문 내용과 방식이 그대로, 혹은 수정보완되어 널리 활용되고 있다(Huang & Cornell, 2015).

그러나 OBQ가 괴롭힘 경험률 조사에서 널리 활용되고 있음에도 불구하고 이러한 질문지를 통해 측정된 경험률이 연구에 따라 크게 차이가 난다는 분석 결과가 거듭 제시되면서 이러한 차이를 초래하는 편향요인들이 괴롭힘 평가 분야의 주요 쟁점이 되어왔다(Jia & Mikami, 2018; Menesini & Salmivalli, 2017; Solberg & Olweus, 2003; Swearer, Siebecker,

Johnsen-Frerichs, & Wang, 2010). 예컨대, Swearer 등(2010)은 괴롭힘 경험률에 대한 선행연구들을 개관하면서 연구에 따라 가해율이 4.8%~26%, 피해율이 5.7%~39%까지 매우 다양하게 보고되고 있음에 주목하였다. 이들은 이러한 큰 차이가 표본 집단의 특성, 문화, 국가 등 배경요소의 차이로 설명할 수 있는 수준을 넘어서는 것으로, 이러한 비일관된 결과는 괴롭힘을 조작적으로 정의하는 방식, 즉 구체적인 측정 방법이나 전략이 연구에 따라 크게 차이가 있기 때문이라고 분석하였다. 유사하게, Jia와 Mikami(2018)도 괴롭힘 측정에 관한 쟁점을 정리한 최근의 개관연구에서 측정의 문제는 괴롭힘 연구에서 “아킬레스 건”(Cornell, Sheras, & Cole, 2006, p. 193)이라는 주장을 인용하면서, 경험률에 있어서 나타나는 이와 같은 큰 차이는 무엇보다도 괴롭힘의 핵심요소를 조작화하는 방식이 연구들 간에 상이하기 때문이라고 주장하였다. 구체적으로 이들은 괴롭힘의 자기보고식 평가에서 의도성이나 힘의 불균형 같은 요소들이 포함된 ‘괴롭힘’ 경험을 측정하느냐, 아니면 이러한 요소들에 대한 고려 없이 행동의 형태에만 기초하여 일반적인 또래대상의 공격행동의 빈도나 정도를 측정하느냐에 따라 경험률이 크게 다른 점을 지적하였으며, 따라서 이 둘을 구분하여 측정하는 방법론적 개선이 필요하다고 역설하였다.

이처럼 핵심 구성요소의 조작화 방식의 차이가 괴롭힘의 자기보고식 평가 결과의 비일관성 혹은 편향을 설명하는 중요한 요인이라 하더라도, 최근의 개관 연구들(이동형, 2014a; Cornell & Cole, 2012; Furlong, Sharkey, Felix, Tanigawa, & Green, 2010; Jia & Mikami, 2018)에 따르면, 이러한 차이 외에도 평가조건이나 절차와 관련된 보다 미묘한 방법론적 차이(예를

들면, 익명 실시, 참조시간 차이, 정의 제시, 문항 제시 순서, 질문의 형태 등)나 응답자의 반응 특성(예를 들면, 무효반응, 예방 프로그램 실시에 따른 민감성 증가, 연구자의 요구특성에 따른 반응 등)을 포함한 보다 다양한 요인들이 괴롭힘 경험의 자기보고 결과를 편향시킬 수 있는 것으로 거론되고 있다. 그러나 이처럼 많은 잠재적 편향요인들이 있다라도, 이러한 요인들의 실제 영향 여부와 구체적인 영향의 정도에 대해서 직접적 검증이 이루어지기 보다는 아직 쟁점 수준에 머물러 있는 경우가 많다. 따라서 이러한 요인들이 실제로 자기보고식 평가 결과에 유의한 영향을 미칠 수 있는지, 미칠 수 있다면 그 정도는 어떠한지 객관적으로 평가하는 실증연구들이 필요하다. 만일 이러한 요인들 중 적어도 일부가 평가 결과를 체계적인 방식으로 편향시킨다면, 이들을 적절히 통제하거나 제거하는 노력이 측정오차와 연구 간의 비일관성을 줄여, 보다 신뢰롭고 타당하며, 정밀한 괴롭힘 평가를 가능하게 할 것이기 때문이다. 특히 평가조건 관련 변인들은 비교적 통제가 용이하며, 괴롭힘 평가 개선을 위해 직접적 시사점을 제공할 수 있으므로, 실증연구가 더욱 절실히 필요한 영역이라고 할 수 있다.

자기보고에 기초한 괴롭힘 경험률 조사 결과를 편향시킬 수 있는 요인들에 대한 우려는 국내의 상황에서도 예외가 될 수 없다. 국내의 경우, 괴롭힘이나 학교폭력에 대한 연구문헌이 지난 10~20년간 급격히 증가한 것에 비하면, 괴롭힘의 측정 및 평가와 관련된 연구는 아직 미미한 실정이다. 현재까지 국내 연구들로는 자기보고식 평가와 또래평가 방법 혹은 교사평가의 차이를 비교한 소수의 연구(신유림, 2006; 정태연, 김인경, 김은정, 2001;

최지영, 2008)가 있으며, 괴롭힘의 자기보고식 평가 관련 쟁점을 정리한 개관연구(이동형, 2014a)와 사회적 바람직성, 적대적 귀인 같은 개인특성 및 익명성, 참조시간을 포함한 평가 조건 변인들의 편향 가능성을 탐색한 한 편의 실증연구(최문영, 이동형, 2015)가 있을 뿐이다.

더욱이, 자기보고식 평가 방법을 활용하고 있는 국내의 대규모 학교폭력 실태조사 결과들을 보면, 조사기관에 따라 보고되는 경험률에서 큰 차이를 보이는 등 괴롭힘 평가에 관한 현 우려와 쟁점들을 그대로 반영하고 있는 것으로 보인다. 가령, 2017년 학교폭력실태 전수조사(교육부, 2017) 결과, 피해율이 0.9%, 가해율이 0.3%로 나타났다. 반면, 비슷한 시기에 푸른나무 청예단(2018)이 전국 6,675명 초중고 재학생을 대상으로 한 실태조사에 의하면, 피해율 6.5%, 가해율 3.9%로 나타났다. 이 두 실태조사가 보고하는 경험률이 단순히 표본조사와 전수조사의 차이로 간주하기에는 과도한 차이를 보인다. 교육부의 보도자료(교육부, 2017)는 학교폭력이 “2012년 이후 하향 안정화 추세”(p. 1)를 보이는 것으로 조사결과를 해석하고 있으나, 괴롭힘 경험률이 조작화 방식이나 미묘한 측정 방법론적 차이, 응답자의 민감성 증가 등 다양한 요인들에 의해 큰 차이를 보일 수 있음을 경고하는 연구문헌을 고려한다면, 이러한 해석은 분명 재평가 및 재논의의 여지가 있는 것이다.

이처럼 괴롭힘의 자기보고식 평가와 관련하여 제기되고 있는 쟁점들과 국내의 연구 현황을 고려할 때, 이러한 쟁점들을 정리할 뿐 아니라 괴롭힘이나 학교폭력 평가 결과의 신뢰도, 타당도, 정밀성 제고를 위해 구체적 시사점을 줄 수 있는 실증연구들이 절실히 필요한 시점임에 틀림이 없다. 이러한 맥락에서 본

연구는 괴롭힘의 자기보고, 특히 경험률 평가에 영향을 미칠 수 있는 요인들 중 객관적 검증이 가능한 세 가지 평가조건 관련 요인들, 구체적으로는 정의제시 여부, 익명실시 여부, 질문형태에 초점을 두어 이러한 요인들이 괴롭힘 경험보고에 미치는 영향을 살펴보고자 한다. 각 요인들과 관련된 선행연구들을 차례로 검토하고 이를 통해 본 연구의 필요성을 구체적으로 제시하면 다음과 같다.

본 연구에서 살펴볼 첫 번째 평가조건 변인은 ‘정의제시’ 여부다. OBQ에서 질문에 답하기에 앞서 괴롭힘에 대한 상세한 정의를 제시하는 것은 괴롭힘의 3대 핵심요소를 조작화하기 위해 오랫동안 채택해 온 방식으로, 이는 무엇이 괴롭힘이고 무엇이 괴롭힘이 아닌지 발달수준에 적절한 정의를 제시한 후, 괴롭힘 경험의 빈도를 표기하도록 하는 방식을 의미한다. Solberg와 Olweus(2003)는 괴롭힘은 학생들이 흔히 사용하는 일상용어인 바, 학술적 의미와 다르게 이해하고 사용할 수 있음을 강조하면서 괴롭힘 실태조사에서 개념 정의를 명확히 제시하는 것이 필수적이라고 주장하였다. 이를 지지하는 증거로, Vaillancourt 등(2010) 연구와 Land(2003)의 연구에서 학생들이 생각하는 괴롭힘이 무엇인지 정의하거나 전형적인 괴롭힘 사례를 제시하도록 요구하였을 때, 힘의 불균형 요소나 반복성의 요소를 포함하여 괴롭힘의 정의를 이해는 경우가 드문 것으로 나타났다. 더 나아가 괴롭힘이라는 용어와 함께, 괴롭힘에 대한 Olweus식의 정의를 제시받은 학생들의 경우, 그렇지 않은 경우보다 피해경험을 유의하게 적게 보고하며(Vaillancourt et al., 2010), 가해경험 또한 적게 보고한다는 다른 연구결과(Kert, Coddling, Tryon & Shiyko, 2009)도 있다.

그러나 정의를 단순히 제시하는 것이 경험을 보고할 확률을 낮추는지에 대해서는 의문의 여지가 있는데, 이는 Huang와 Cornell(2015)의 연구에서 정의제시 조건과 비제시 조건 간에 경험률에서 유의한 차이가 없는 것으로 나타났다기 때문이다. 즉 정의를 제시하더라도 학생들이 이에 충분히 주의를 기울이지 않는다면, 그들의 괴롭힘 보고에 아무런 영향도 미치지 않을 수 있다. 이러한 가능성에 대해 Hamby와 Finkelhor (2000)는 괴롭힘이라는 용어를 사용하는 것 그 자체가 학생들에게 정서적 부담감을 줄 수 있으며, Furlong 등(2010)도 학생 스스로 괴롭힘 피해 사실을 시인하는 것이 개인적 약점을 시인한다는 면에서 부담이 될 뿐 아니라, 피해자라는 자기상을 내면화하지 못하고 이러한 정의와 일치하지 않는 방식으로 응답할 가능성이 있다고 주장하였다. 흥미롭게도, 보다 정교하게 설계된 Baly와 Cornell(2011)의 연구에서는, 괴롭힘 정의를 제시하는 OBQ를 실시하되 실험집단에 무선배정된 중학생들은 괴롭힘의 힘의 불균형 요소를 설명하는 짧은 교육용 영상을 시청한 후 OBQ에 응답하게 하였고, 통제집단의 중학생들은 이러한 절차 없이 OBQ에 응답하도록 하였을 때, 실험조건의 청소년들이 무려 48~54%나 적게 괴롭힘 경험을 보고한 것으로 나타났다.

이러한 일련의 실증연구 결과들은, 괴롭힘에 대한 ‘정의제시’가 경험보고에 영향을 미칠 수 있지만, 단지 기계적으로 정의를 제시하는 것 이상의 (제시된 정의를 학생들이 읽고 이해하도록 돕는) 추가 조치가 있을 때, 자기보고 측정치의 신뢰도를 높일 수 있음을 시사한다. 우리나라의 경우, 괴롭힘이라는 용어가 학생들에게 어떤 의미인지에 대한 국내 학생들 대상의 기초 연구가 필요한 상황이다. 그러나

13개 이상의 언어로 번역된 영어의 bullying이 언어에 따라 포괄하는 행동의 범위가 상당히 다르다는 선행연구 결과(Smith, Cowie, Oafsson, & Liefoghe, 2002)를 참조할 때, 국내에서도 괴롭힘이라는 용어의 학문적 의미와 학생들이 이해하는 개념이 차이가 있을 가능성은 매우 높으며, 따라서 괴롭힘에 대한 정의제시가 경험률 평가에 어떠한 영향을 미치는지 살펴보는 국내 실증연구가 필요하지만, 이러한 연구는 아직 없었다. 이에 본 연구에서는 정의제시 여부가 괴롭힘 경험의 자기보고에 미치는 영향을 검증하기 위해, 괴롭힘에 대한 OBQ 방식의 정의를 제시하되, 이러한 정의를 읽었는지 여부를 확인하고 반드시 이러한 정의를 읽은 경우에만 괴롭힘 관련 경험여부를 보고하도록 하는 방식의 추가 조치를 취한 후, 정의제시 조건에서 경험률이 유의하게 낮은지 검증하고자 한다(연구방법 참조).

본 연구에서 그 영향을 검증할 두 번째 평가조건 변인은 ‘익명실시’ 여부다. 괴롭힘 경험률 조사는 ‘솔직한 응답을 촉진하기 위해’ 흔히 익명으로 실시되며, 또 이러한 방법이 권장되기도 한다(Solberg & Olweus, 2003). 아마도 비슷한 이유에서, 그리고 질문 내용의 민감성을 고려하여, 교육부나 청예단의 학교폭력 실태조사도 현재 모두 익명으로 실시되고 있다. 특히 OBQ는 원래 익명 실시용 도구로 개발되었기 때문에 이러한 취지에 따라 대부분의 조사연구에서 익명으로 실시되어왔다. 그러나 익명실시의 경우, 결과의 타당성을 외적인 준거를 통해서 확인하기 어렵다는 중요한 한계점이 있으며, 이로 인해 이러한 실시 방식에 대해 재고할 것을 촉구하는 학자들도 있다(Chan, Myron, & Crawshaw, 2005; Cornell & Bandyopadhyay, 2010; Cornell & Cole, 2012). 이

들에 따르면 익명실시 대신 ‘비밀보장을 전제로 한 적절한 기명실시 절차’가 확보된다면 타당성 확인이나 평가정보의 활용도 측면에서 보다 나은 대안이 될 수 있다는 것이다.

익명실시의 효과를 실증적으로 검증한 연구는 아직 소수에 불과하지만, 이러한 연구들은 익명실시와 기명실시 조건 간에 가해율 및 피해율에서 유의한 차이가 없거나(Chan et al., 2005), 가해, 피해뿐 아니라 동조, 방관, 방어 경험의 ‘정도’를 나타내는 괴롭힘 참여역할의 연속 측정치(리커트형 척도)에서도 유의한 차이가 없거나, 통계적으로는 유의하더라도 실제적으로는 큰 의미가 없는 미미한 차이만을 보고하고 있다(최문영, 이동형, 2015). 또한 괴롭힘에 대한 연구는 아니지만 약물사용, 불법적 행동 등 다른 민감한 행동의 경험률 조사에서도 비밀보장이 전제된다면 기명실시가 반드시 과소보고를 유도하지는 않음을 보여주는 다른 증거(O'Malley, Johnston, Bachman, & Schulenberg, 2000)도 있다. 따라서 괴롭힘의 자기보고식 평가에서 익명실시가 응답자로 하여금 더 솔직하게 응답하도록 돕는다는 기본가정은 실증적으로 검증될 필요가 있으며, 이를 검증한 소수의 연구들이 있었지만, Chan 등(2005)의 연구도, 최문영과 이동형(2015)의 연구도 OBQ와는 정의제시나 측정방식에서 상당한 차이가 있는 평가도구들을 활용하였다는 점에서 이를 확인하는 추가 연구가 필요하다고 할 수 있다. 이에 본 연구에서는 OBQ를 사용하여 익명실시 조건과 ‘비밀보장을 전제로 한’ 기명실시 조건 간에 OBQ를 통해 보고되는 경험률이 유의하게 차이가 있는지를 검증할 것이며, 더 나아가 두 조건에서 솔직히 응답한 정도나, 비밀보장을 확신하는 정도에 대해서도 추가로 질문하여 두 실시조건 간에

괴롭힘 보고의 심리적 경험이 차이가 있는지를 살펴볼 것이다.

세 번째로, 본 연구에서는 정의제시 및 익명실시 여부와 함께 ‘질문형태’(단일문항의 전반질문 대 복수문항의 세부질문)에 따라 괴롭힘 경험률이 다르게 나타나는지 살펴보고자 한다. 괴롭힘 관련 설문조사에서 단일문항을 통해 전반적으로 묻는 형태(즉, 하나의 잘 정의된 문항으로 전반적인 괴롭힘 경험 여부나 빈도를 묻는 방식)와 복수의 문항을 통해 세부적으로 묻는 형태(즉, 괴롭힘을 다양한 하위 유형으로 구분하여 여러 문항을 통해서 경험 여부나 빈도를 묻는 방식) 중 어느 방식이 보다 적절한지는 괴롭힘의 자기보고식 평가에서 쟁점 중 하나이다(이동형, 2014a; Cornell & Cole, 2012). 일부 학자들은 전반질문은 신뢰도가 낮을 가능성이 있으므로 복수문항을 사용한 심리측정적 접근을 통해 괴롭힘 경험을 측정해야 한다고 주장한다(Sharkey, Dowdy, Twyford, & Furlong, 2012). 반면, 괴롭힘 외에도 다수의 영역에 대해서 묻는 대규모 실태조사 등에서는 단일문항의 전반질문을 통해 괴롭힘 경험여부를 간편하게 파악할 수 있으며, 실태조사 결과의 해석, 비교, 활용 또한 수월한 장점이 있으므로(Solberg & Olweus, 2003), 이러한 형태의 질문이 보다 널리 활용되고 있다. 사실, OBQ에서는 단일문항의 전반질문을 먼저 제시한 후, 이어 구체적인 세부유형별 경험을 다문항 형태로 묻도록 구성되어 있어서 두 방식이 병용되고 있다. 또한 선행연구(Solberg & Olweus, 2003)에서 이 두 질문형태로 측정된 경험보고 간의 상관은 .77~.79로 나타나 두 질문형태에 걸쳐 학생들이 대체로는 일관되게 반응함이 시사되었다. 그러나 실제 괴롭힘 경험률 평가 시, 질문형태에 따라 경험률 측정

치가 얼마나 다르게 나타나는지, 만일 OBQ의 방식처럼 두 질문형태가 병용될 필요가 있다면 그 실증적 근거는 무엇인지 아직 명확히 이해되고 있지 못하다.

질문형태와 관련된 실증 연구로, OBQ에서 질문형태에 따른 경험률의 차이를 분석한 Vaillancourt 등(2010)의 연구가 있다. 이 연구에서 OBQ의 전반질문을 세부질문과 대비하여 괴롭힘 경험 유무를 분류할 때 민감도와 특이도가 어떠한지 평가하였는데, 전반질문은 괴롭힘 경험이 실제로 없는 학생들을 정확히 가려내는 정도, 즉 특이도(specificity)는 상대적으로 높지만, 그러한 경험이 있는 학생들을 정확히 찾아내는 정도, 즉 민감도(sensitivity)는 높지 않은 것으로 나타났다. 국내외의 대규모 설문조사에서 괴롭힘 경험을 측정하기 위해 단일문항의 전반질문을 사용하는 경우가 흔하고, 교육부나 청예단의 학교폭력 실태조사에서 전반질문과 세부질문의 두 형태가 병용되고 있는 점, 분류 정확도(diagnostic accuracy) 등 전반 대 세부질문의 구체적 기능을 살펴본 실증연구가 아직 국제적으로도 부족한 점을 고려한다면, Vaillancourt 등(2010)의 연구결과를 확인하거나 확장하는 국내 연구는 더욱 의미가 있을 것으로 보인다. 이에 본 연구에서는 질문형태에 따라 가해율 및 피해율이 유의한 차이를 보이는지, 그리고 분류의 정확도를 반영하는 여러 지표들, 즉 질문형태에 따른 일치율, 증감률, 상관, 민감도와 특이도 등을 살펴봄으로써, 괴롭힘이나 학교폭력 실태조사에서 흔히 사용되는 이러한 두 질문형태의 차별적 기능을 살펴보고자 한다.

본 연구에서는 이상으로 살펴본 3가지 평가조건 변인이 괴롭힘 경험의 자기보고를 체계적으로 편향시킬 가능성을 알아보기 이들 변

인에 따른 경험률의 차이를 비교하되, 절단점을 고려하여 분석하고자 한다. 괴롭힘을 정의하는 핵심요소 중 하나는 ‘반복성’인데, OBQ는 ‘지난 2~3개월’ 동안의 괴롭힘 가해 및 피해경험을 학생이 회상해 내도록 요구한다. 이때 괴롭힘의 반복성 요건을 충족하기 위해서 필요한 최소 경험 빈도, 즉 절단점(cut-off points)을 어떻게 설정하느냐에 따라 괴롭힘 경험률 측정치는 차이를 가져올 수밖에 없다. 그러나 괴롭힘 경험을 측정하는 도구와 연구자에 따라, ‘가끔’ 혹은 ‘자주’ 등과 같은 모호한 절단점을 사용하는 경우도 있고, ‘한두 번’이나 ‘1주에 1회 이상’ 등과 같이 다양한 빈도값이 반복성을 조작화하는데 사용되고 있다 (Swearer et al., 2010). OBQ의 경우, 지난 2~3개월 동안 ‘1~2회만’ ‘월2~3회’ ‘주1회’ ‘1주일에 여러 번’의 네 가지 절단점 사용이 가능하다. 어의적으로 보면 ‘1~2회만’도 반복적 경험으로 간주할 수는 있겠지만, Solberg와 Olweus (2003)는 ‘월2~3회’를 가장 ‘합리적이고 유용한’ 절단점으로 권장한 바 있다. 그 이유는 이 절단점에 기초하여 경험 유무를 결정하였을 때 우울, 자존감, 또래수용 등에서 경험집단과 무경험 집단 간에 차이가 있는 것으로 나타났기 때문이다.

본 연구에서는 이러한 권고에 따라 ‘월2~3회’를 절단점(‘엄격한’ 절단점)으로 적용하여 괴롭힘 경험률을 분석하고 그 결과를 주로 해석할 것이지만, 보다 낮은 수준(‘1~2회만’)의 절단점(즉 ‘관대한’ 절단점)을 적용하여 산출한 경험률도 함께 비교 제시할 것이다. 이는 낮은 절단점을 적용하여 경험률을 보고하는 연구들도 있을 뿐만 아니라 이렇게 두 절단점에 따른 결과를 비교함으로써, 반복성/지속성이 있으며 임상적으로 보다 의미있는 괴롭힘 경

험률(엄격한 절단점을 적용할 경우)과 괴롭힘 뿐 아니라 산발적이거나 비교적 경미한 형태의 폭력 경험률(관대한 절단점을 적용할 경우)을 괴롭힘에 포함시킬 때 나타나는 경험률 증감 정도나 절단점에 따른 평가조건 변인의 영향의 차이 등을 살펴볼 수 있기 때문이다.

끝으로, 본 연구에서는 평가조건 변인 및 절단점에 따른 경험률 분석과 함께 성별에 따른 분석 결과를 제시하고자 한다. 국내외의 괴롭힘이나 학교폭력 실태조사 연구에서 성별에 따른 차이는 대체로 남학생의 경험률이 높은 것으로 나타나고 있지만, 직접적 괴롭힘은 남학생의 경험률이 보다 빈번한 반면, 관계적 괴롭힘 같은 간접적 형태는 성차가 없거나 여학생이 더 빈번히 경험하는 것으로 보고하는 선행연구들도 있다(조운오, 2013; 이은주, 2001; Crick & Grotpeter, 1995). 괴롭힘 연구에서 흔히 보고되는 이러한 성차의 패턴은 경험률 연구에서 성차를 고려한 분석이 필수적임을 시사한다. 그러나 편향요인을 탐색하는 본 연구에서 절단점과 함께 성별을 고려하여 분석하는 다른 이유는, 유의한 성차가 실제 성차를 반영할 수도 있지만, 성에 따른 자기보고 양상의 차이로 인해 초래된 편향을 반영할 수도 있다는 주장(Rutter, Caspi, & Moffitt, 2003)이 있기 때문이다. 예컨대, 남학생들은 여학생들에 비해 괴롭힘의 전반적 경험을, 혹은 특정한 유형의 괴롭힘 경험을 보다 잘 회상하거나 기꺼이 보고할 가능성이 있다. 즉 성별은 내재화 문제의 자기보고식 평가에서 여학생들이 남학생들에 비해 보다 빈번히, 그리고 기꺼이 부정적 정서를 보고하는 경향이 있는 것과 유사한 맥락에서 괴롭힘 보고의 편향요인으로 작용할 가능성을 배제할 수 없다는 점이다. 성별이 편향요인으로 작용하는지의 여부는 본

연구의 주 관심사가 아니며 연구 설계 상 이를 직접 검증하지는 않을 것이다. 다만 서로 다른 절단점에 따라 괴롭힘 경험률의 성차가 일관되게 나타나며 그 크기가 어떠한지를 검토함으로써 성별이 편향요인으로 작용할 가능성을 가늠해 보고자 한다.

이상으로 제시한 선행연구들과 연구의 필요성에 기초하여, 본 연구의 목적은 자기보고식 괴롭힘 경험률 평가의 잠재적 편향요인으로 쟁점이 되고 있는 여러 요인들 중 평가조건 관련 변인, 구체적으로는 익명실시 여부, 정의 제시 여부, 질문형태에 따라 남녀 중학생이 보고하는 괴롭힘 가해 및 피해율이 다르게 나타나는지를 서로 다른 절단점을 적용하여 검증함으로써 이러한 변인들이 경험률 평가결과를 편향시킬 가능성을 살펴보는 것이다. 이를 위해 설문지 지시문을 통해 두 가지 평가조건 변인(익명실시/기명실시, 정의제시/정의미제시)을 달리한 4개의 평가조건에 연구대상 남녀 중학생들을 무선배정한 후, 성별에 따른 차이, 각 평가조건 변인에 따른 차이, OBQ의 질문형태에 따른 차이를 두 가지 절단점을 고려하여 비교해 볼 것이다. 본 연구에서 살펴 볼 구체적인 연구문제를 제시하면 다음과 같다.

연구문제 1. 성별과 절단점에 따라 중학생의 괴롭힘 경험률(가해율 및 피해율)이 차이가 있는가?

연구문제 2. 정의제시 여부(정의제시 대 미제시)와 절단점에 따라 중학생의 괴롭힘 경험률이 차이가 있는가?

연구문제 3. 익명실시 여부(익명실시 대 기명실시)와 절단점에 따라 중학생의 괴롭힘 경험률이 차이가 있는가?

연구문제 4. 질문형태(전반질문 대 세부질

문)와 절단점에 따른 중학생의 괴롭힘 경험률의 차이와 분류 정확도는 어떠한가?

방 법

연구대상

경북, 경남, 부산, 울산에 소재한 7개 중학교 1~3학년에 재학 중인 남녀 중학생 725명을 대상으로 설문조사를 실시하였다. 연구에 참여할 학교와 담당교사를 섭외를 하면서 연구의 목적, 협조사항 등을 안내하였고, 담당교사는 연구대상자인 학생들에게 연구목적을 설명하고 학생이 참여에 동의하는 경우 설문지를 작성하도록 안내하였다. 설문조사는 교사가 직접 감독하는 가운데 이루어졌으며, 학급 단위로 이루어졌으며, 실시 직후 그 자리에서 회수하는 방식으로 자료를 수집하였다. 회수된 725명의 자료 중 불성실하게 응답한 8부를 제외하였다. 또한 본 연구의 목적상 ‘기명 실시’ 조건의 설문지를 받은 대상자들 중 개인 인적사항(반, 번호)을 기입하지 않았거나, ‘정의 제시’ 조건에서 정의를 읽었는지를 확인하는 질문에 답변하지 않았거나 읽지 않았다고 응답한 설문지 27부도 통계분석에서 제외시켜, 최종적으로 690명의 자료가 통계분석에 포함되었다. 연구대상자의 성별, 학년별 분포를 보면, 여학생이 336명(1학년 143명, 2학년 45명, 3학년 148명), 남학생은 354명(1학년 63명, 2학년 22명, 3학년 269명)이었다.

연구절차

본 연구에서 전체 연구대상자들은 두 가지

평가조건 변인인 정의제시 여부(즉, 정의제시 혹은 정의미제시) 및 익명실시 여부(즉, 익명 실시 혹은 기명실시)가 결합된 네 조건 중 하나에 무선 배정하였다. 네 조건에 대응하는 네 유형의 질문지가 개발되었는데, 이들은 지시문 내용과 정의 제시 여부에서만 차이가 있을 뿐, 그 외에는 모두 동일한 내용의 설문지였다(평가조건 변인의 구체적인 조작 내용은 다음 절 참조). 무선배정의 정의 상 각 대상자가 네 유형의 질문지를 제시받을 확률이 동일해야 하므로, 본 연구에서는 다음과 같은 절차에 따라 무선 배정하였다. 먼저 각 조건별로 동일한 부수의 설문지를 인쇄하여 유형별로 순서(즉, 유형1, 유형2, 유형3, 유형4, 유형1, 유형2...순)대로 배열한 후, 성별, 학년, 소속 학급, 학교, 다른 평가조건 등, 괴롭힘 경험률에 체계적 영향을 미칠 수 있는 다른 변인들과 무관하게 모든 연구대상자에게 네 유형의 설문지를 무선적으로 배포하였다. 이러한 무선 배정 절차를 거쳐 설문조사를 실시한 후, 최종적으로 유효한 자료로서 통계분석에 포함된 대상자 수는 총690명이었으며, 이를 평가조건 별로 분류하면, 정의제시/익명실시 조건이 177명, 정의제시/기명실시 조건이 173명, 정의미제시/익명실시 조건이 169명, 정의미제시/기명실시 조건이 171명이었다. 본 연구에서 검사 설문지는 검사 전체 지시문, OBQ의 표준적인 괴롭힘 정의 제시, OBQ의 가해경험 문항, OBQ의 피해경험 문항, 사후질문 순으로 구성되었다.

평가조건 변인의 조작

본 연구에서 고려한 평가조건 변인은 정의 제시 여부, 익명실시 여부, 질문형태이다. 질문형태 면에서 OBQ는 전반질문을 먼저 제시

하고 연이어 세부질문을 제시하는 형식을 취하고 있으므로, 모든 대상자들이 이 두 형태의 질문에 모두 답변하게 된다. 즉 이 변인의 경우 연구설계를 통해 별도의 조작이 필요하지 않은 반복측정 요인이었다.

정의제시와 익명실시 여부는 각각 설문지의 전체 지시문과 OBQ의 표준적인 정의제시 방식을 통해 조작되었다. 정의제시 조건의 설문지는 원래 OBQ의 방식대로 괴롭힘에 대한 정의를 먼저 제시한 후 괴롭힘 가해 및 피해를 묻는 문항을 제시하였고, 정의미제시 조건은 이러한 정의를 생략한 채 바로 괴롭힘 경험을 묻는 문항을 제시하였다. 본 연구에서는 OBQ에서 제시하는 표준적인 정의를 그대로 번역하여 제시하되, 학생들이 쉽게 간과하기 쉬운 괴롭힘의 핵심 구성요소를 상기시키는 그림과 간단한 설명을 추가로 삽입하여 괴롭힘과 괴롭힘이 아닌 것을 구분하기가 보다 쉽도록 제시하였다(그림 1 참조). 또한, 제시된 정의를 제대로 '읽고' 이해한 후에 OBQ의 질문에 답하도록 유도하기 위해, "반드시 읽어주세요"라는 지시문과 함께 그림 1의 정의를 제시하였고, 이후 괴롭힘 경험여부를 묻는 첫 질문을 제시하기에 앞서 "제시된 괴롭힘의 정의를 다 읽어보았나요?"라는 질문에 예 혹은 아니요로 응답하도록 하였다. 이때 아직 읽지 않은 경우는 "다시 돌아가서 괴롭힘의 정의를 반드시 읽어주세요"로 재지시하여 정의를 반드시 읽고 응답하도록 유도하였다. 또한 OBQ의 전체 문항을 다 완성한 후 사후질문 중 하나로 "앞에 제시된 괴롭힘 정의를 다 읽고 나서 설문에 응답했나요?"로 질문하여 정의를 읽고 응답하였는지 재점검하였으며, 이 때 아니요로 응답하거나, 이 질문에 답하지 않았다면 제시된 정의를 읽지 않은 것으로 간주하여 통계분

석에서 제외시켰다.

익명실시 여부는 전체 지시문을 통해 조작되었는데, 익명조건의 설문지에는 "이 설문지에 여러분이 응답하는 내용은 연구목적으로만 사용되므로, 다른 친구들이나 선생님들에게 절대로 알려지지 않습니다. 따라서 학년과 성별만 기입해 주십시오."라는 지시문을 강조하여 제시한 반면, 기명조건의 설문지는 "여러분이 응답한 내용은 다른 친구들에게 절대로 알려지지 않지만, 여러분의 학교생활을 알아보기 위해 선생님께서 여러분이 응답한 내용을 확인하실 수 있습니다. 따라서 성별, 학년, 반, 번호를 꼭 적어주세요."라는 지시문을 제시함으로써, 익명조건과 마찬가지로 비밀보장을 전제하였지만, 필요한 경우 교사의 권한 하에 응답자 개인을 식별하고 그 내용을 확인할 수 있음을 알려주었다. 기명조건에 배정된 연구대상자 중 개인 확인정보(반, 번호)를 기입하지 않은 경우는 기명조건에서 응답하였다고 간주하기 어려우므로 분석에서 제외시켰다.

연구 도구

Olweus 괴롭힘 척도

괴롭힘 경험률, 즉 가해율과 피해율을 측정하기 위해 2007년 개정된 Olweus Bullying Questionnaire(OBQ: Olweus, 2007)를 이동형(2014b)이 국내의 실정에 맞게 수정·보완한 것 중 괴롭힘에 대한 정의 제시 부분과 가해경험과 피해경험을 묻는 문항만을 발췌하여 사용하였다. 이 척도는 '지난 2~3개월' 동안 학교에서 내가 다른 아이를 괴롭힌 일(가해경험)과 다른 아이가 나를 괴롭힌 일(피해경험)을 각각 10개의 문항을 통해 묻도록 구성되어있다. 가해경험을 묻는 10문항의 경우, 첫 번째 문항

은 전반질문(“지난 2~3개월 동안 내가 다른 학생(한명 혹은 여러 명)을 괴롭힌 일이 얼마나 자주 있었나요?”)이며, 나머지 9문항은 다양한 유형의 괴롭힘 별로 구체적인 가해경험을 묻는 9개의 세부질문으로, 신체폭력, 언어폭력, 금품갈취, 강요·협박, 따돌림, 성폭력, 사이버폭력 경험을 묻는다. 연이어 제시되는 피해경험을 묻는 10문항도 가해경험과 동일한 방식으로 전반질문 1문항과 세부질문 9문항으로 구성되어 있다.

OBQ에서 가해 및 피해경험을 묻는 각 질문에 대해서, ‘전혀 없음’, ‘1-2번만’, ‘월2-3회’, ‘주1회’, ‘1주일에 여러 번’ 중 하나를 선택하여 응답하도록 되어있다. 이러한 5가지 응답 중, ‘전혀 없음’을 제외한 나머지 4가지 선택지를 괴롭힘 경험의 유무를 결정하는 절단점

으로 사용할 수 있는데, 본 연구에서 ‘관대한 절단점’을 사용한 경우에는 ‘1-2번만’이나 그 이상의 빈도를 선택한 경우에만 괴롭힘 경험이 있는 것으로 간주하여 재코딩하여 경험률을 산출하였고, ‘엄격한 절단점’을 사용한 경우에는 ‘월2-3회’나 ‘주1회’, 혹은 ‘1주일에 여러 번’을 선택한 경우에만 괴롭힘 경험이 있는 것으로 보고 경험률을 산출하였다.

한편, OBQ의 전반질문은 단일문항이므로, 이러한 두 절단점을 적용하면 쉽게 가해 및 피해 경험여부가 결정될 수 있는 반면, 세부질문의 경우에는 가해경험 및 피해경험 별로 각각 9개의 문항으로 구성되어 있으므로, 이 9개의 문항 중 몇 개 이상의 문항에서 절단점보다 높은 빈도를 나타내면 괴롭힘 경험이 있는 것으로 결정할지 또 하나의 기준이 필요

★ ‘괴롭힘’이란? ★

어떤 학생(들)이 다른 학생에게 다음과 같은 일을 할 때 ‘괴롭힘’이라고 합니다.

- 상스럽고 기분 나쁜 말을 하거나, 놀리거나, 욕을 한다.
- 완전히 무시하거나, 따돌리거나, 어떤 일을 할 때 일부러 끼워주지 않는다.
- 때리거나, 발로 차거나, 떠밀거나, 어떤 곳에 밀어 넣고 못 나오게 한다.
- 거짓말을 하거나, 나쁜 소문을 퍼트리거나, 쪽지를 보내거나 하여 다른 아이들이 싫어하게끔 만든다.
- 이 외에 다른 방식으로 상처를 주는 일을 한다.

<p>괴롭힘인 경우</p> <p>위와 같은 일이 반복해서 일어남</p> <p>당하는 학생이 스스로 방어하기 어려움</p>	<p>괴롭힘이 아닌 경우</p> <p>친한 친구 사이의 장난</p> <p>비슷한 힘을 가진 친구끼리 싸우는 것</p>
------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------



*한쪽이 방어가 어려운 경우



*비슷한 힘 가진 친구

그림 1. 괴롭힘의 정의 제시

하다. 본 연구에서는 OBQ에 대한 선행연구(Solberg & Olweus, 2003)에 기초하여, 세부질문 9문항 중 적어도 1문항 이상에서 절단점 이상의 빈도를 보고한 경우는, 괴롭힘 경험이 있는 것으로 간주하여 경험률을 산출하였다.

사후 질문

본 연구에서는 OBQ의 가해 및 피해경험을 묻는 20문항을 실시한 후 연구자가 제작한 3개의 사후질문에 대해 추가로 답변하도록 하였다. 첫 번째 질문은 정의제시 조건에서 정의를 읽고 응답하였는지를 묻는 문항으로, 정의가 제시된 설문지의 경우에만 포함이 되었다(앞 절의 ‘평가조건 변인의 조작’ 참조). 두 번째 질문은 연구대상자가 가해 및 피해경험 질문에 대해 솔직하게 응답한 정도를 5점 척도(매우 솔직하게~매우 솔직하지 않게) 상에 평정하도록 요구하는 문항이었다. 세 번째 질문은 비밀보장을 확신하는 정도, 즉 자신이 응답한 내용이 다른 친구들에게 알려질 것이라고 생각하는 정도 5점 척도(절대 알려지지 않음 ~ 틀림없이 알려질 것임) 상에 각각 평정하도록 구성하였다. 첫 질문을 제외하고 두 번째 및 세 번째 질문은 익명실시 여부에 따라 응답자들의 심리적 경험(솔직성, 비밀보장의 확신 정도)이 차이가 있는지를 탐색해 보기 위한 참고 질문이었다.

자료 분석

성별 및 평가조건 변인에 따라 중학생의 괴롭힘 경험률(경험빈도)이 다른지를 알아보기 위해, 먼저 괴롭힘 가해경험 및 피해경험을 보고한 빈도(사례수)와 백분율을 산출하고, 평가조건 변인들에 따라 이러한 빈도가 통계적

으로 유의하게 차이가 있는지 교차분석을 통한 독립성 검증(χ^2 검증)을 실시하였다. 가해 및 피해경험 유무는 2개의 전반문항과 18개의 세부분항 별로 각각 나누어서 살펴보고, 관대한 절단점(‘1~2회만’)과 엄격한 절단점(‘월 2-3회 이상’)을 모두 적용하여 분석하였다. 다만 엄격한 절단점을 적용할 경우, 기대빈도 값이 5이하인 셀이 20% 미만이어야 한다는 χ^2 검증의 가정에 위배되는 경우(즉, 네 셀 중 적어도 한 셀에서 기대빈도 값이 5이하인 경우)가 발생한다면, Siegel과 Castellan(1988)의 권고에 따라 χ^2 값의 유의성 검증을 위해 Fisher의 정확률검증을 실시하였다. 사후질문에 대한 리커트 척도 상의 점수의 평균차이를 검증하기 위해서는 독립표본 t 검증을 사용하였다.

한편, 질문형태에 따른 경험률의 차이는 이분변수의 반복측정치에 대한 분석이므로 χ^2 값의 유의성 검증을 위해 이항분포를 사용한 McNemar 검증을 적용하였다. 또한 질문형태에 따른 분류의 정확도를 평가하기 위해 질문형태에 따른 증감률, 일치율(correspondence rate)과 이분변수 간 Pearson 상관계수(파이계수), 전반문항의 민감도와 특이도를 산출하였다.

본 연구에서 일치율은 진긍정률(true positive rate)과 진부정률(true negative rate)을 합산한 값으로, 전체 대상자 중 전반질문과 세부질문 모두에 대해 괴롭힘 경험유무를 일치하는 방식으로 보고한 대상자의 비율을 의미한다. 일치율의 경우 표본에서 관찰빈도가 매우 낮은 경험의 유무를 평가할 경우(즉, 엄격한 절단점 적용할 경우), 진부정 사례가 매우 높게 집계되므로, 두 질문형태 간의 일관성이나 분류 정확도의 지표로 해석하는 데는 한계가 있다. 따라서 이분변수 간의 상관인 파이계수와 전반질문의 민감도와 특이도를 산출하였다.

전반질문의 민감도는 진긍정률을 진긍정률과 위부정률(false negative rate)의 합으로 나눈 값으로, 세부질문에서 괴롭힘 경험이 있는 것으로 보고한 대상자 중 전반질문에 의해 그러한 경험이 실제로 있는 것으로 확인된 대상자의 비율을 의미한다. 반면, 전반질문의 특이도는 진부정률을 진부정률과 위긍정률(false positive rate)의 합으로 나눈 값으로, 세부질문에서 괴롭힘 경험이 없는 것으로 보고한 대상자 중 전반질문에 의해 그러한 경험이 실제로 없는 것으로 확인된 대상자의 비율을 의미한다. 즉 민감도는 괴롭힘 경험이 실제로 있는 대상자를 정확히 가려내는 정도, 특이도는 괴롭힘 경험이 실제로 없는 대상자를 정확히 가려내는 정도를 의미하는 지표라 할 수 있다. 본 연구에서 모든 통계분석은 SPSS 22.0을 활용하였다.

결 과

성별과 절단점에 따른 괴롭힘 경험률의 차이

전체 연구대상자 690명 중 성별에 따른 괴롭힘 가해율과 피해율을 절단점을 달리하여 분석한 결과가 표 1에 제시되어있다. 전반질문에 따른 가해율과 피해율, 세부질문에 따른 가해율과 피해율 순으로 주요 결과를 살펴보면 다음과 같다.

전반질문에 대한 응답을 기초로, 관대한 절단점을 사용하여 집계한 남녀 전체의 가해율은 13.8%, 엄격한 절단점을 사용하여 집계한 남녀 전체의 가해율은 4.1%였다. 성별에 따른 차이를 살펴보면, 관대한 절단점을 사용하였을 경우, 가해율은 남학생(21.2%)이 여학생(6%)

보다 높은 것으로 나타났으며, 엄격한 절단점을 사용하였을 경우의 가해율 또한 남학생(7.3%)이 여학생(0.6%)보다 높게 나타났다. 성별에 따른 이러한 차이는 교차분석 결과(표 1 참조), 모두 통계적으로 유의하였으며(관대한 절단점의 경우 $\chi^2=33.7, p<.001, \phi=-.22$; 엄격한 절단점의 경우 $\chi^2=20.17, p<.001, \phi=-.17$), 성별과 가해율 간의 관련성은 각각 $\phi^2=.048$ 과 $\phi^2=.028$ 로 가해경험 변량의 4.8%(관대한 절단점)와 2.8%(엄격한 절단점)가 성별에 의해 설명되는 것으로 볼 수 있다.

또한 관대한 절단점을 사용하여 집계한 남녀 전체의 피해율은 11.3%, 엄격한 절단점을 사용한 경우에는 4.5%로 나타났다. 성별에 따른 차이는 관대한 절단점과 엄격한 절단점을 사용한 경우 각각, 남학생 16.1% 대 여학생 6.3%, 남학생 7.6% 대 여학생 1.2%였다. 이러한 성차는 교차분석 결과(표 1 참조) 모두 통계적으로 유의하였으며(모든 $p < .001$), 성별이 피해경험 변량의 대략 2.6%($\phi=-.16$)을 설명하는 것으로 나타났다.

이러한 결과는 괴롭힘 경험에 산발적 수준의 비교적 경미한 폭력경험을 포함시키든, 반복/지속적이며 임상적으로 보다 의미 있는 수준의 괴롭힘에 한정하든 상관없이, 남학생의 경험률이 여학생보다 일관되게 높음을 보여준다. 성차의 양상은 동일하지만 다만 관대한 절단점을 사용할 경우, 엄격한 절단점을 사용한 경우에 비해, 경험률이 대략 2~3배 이상 높게 나타남을 알 수 있다.

다음으로 세부질문에 따른 주요 결과를 살펴보면 다음과 같다. 괴롭힘의 구체적인 유형별로 괴롭힘 가해율을 살펴보면, 관대한 절단점을 사용한 경우, 가해율은 1.7%(금품갈취, 강제심부름/강요/협박) ~ 22.5%(욕/놀림/조롱)

표 1. 성별과 집단점에 따른 중학생의 괴롭힘 경험률(가해율 및 피해율)의 차이 분석 결과

분류분류 (분류내용)	가해율						피해율									
	집단점 1~2학년*			집단점 2~3회 이상*			집단점 1~2학년*			집단점 2~3회 이상*						
	성별	남 빈도 (%)	여 빈도 (%)	전체 빈도 (%)	χ^2 (ϕ)	파이 계수	성별	남 빈도 (%)	여 빈도 (%)	전체 빈도 (%)	χ^2 (ϕ)	파이 계수				
전투결과	있음	75 (21.2)	20 (6.0)	95 (13.8)	33.70 (.00)	-.22***	26 (7.3)	2 (0.6)	28 (4.1)	20.17 (.00)	-.17***	27 (7.6)	4 (1.2)	31 (4.5)	16.64 (.00)	-.16***
	없음	279 (78.8)	316 (94.0)	595 (86.2)			328 (92.7)	394 (95.7)	662 (88.7)			327 (92.4)	332 (98.8)	659 (95.5)		
육살림조용	있음	112 (31.6)	45 (12.8)	155 (22.5)	35.13 (.00)	-.25***	25 (6.5)	2 (0.6)	25 (3.6)	17.20 (.00)	-.16***	27 (7.6)	5 (1.5)	32 (4.6)	14.69 (.00)	-.15***
	없음	242 (68.4)	293 (87.2)	535 (77.5)			331 (93.5)	394 (96.4)	665 (86.4)			327 (92.7)	331 (98.5)	658 (95.4)		
무시/따돌림 (가위주지 않음)	있음	35 (9.3)	32 (9.5)	65 (9.4)	0.01 (.95)	.00	2 (0.6)	2 (0.6)	4 (0.6)	0.00 (.96)	.00	18 (5.1)	27 (8.0)	45 (6.5)	2.46 (.12)	-.06
	없음	321 (90.7)	304 (90.5)	625 (90.6)			352 (94.9)	394 (95.4)	686 (95.4)			336 (94.9)	309 (92.0)	645 (95.1)		
매라기, 발본잡기, 머뭇기, 김금 (상체폭력)	있음	35 (9.3)	7 (2.1)	40 (5.8)	16.54 (.00)	-.16***	10 (2.8)	0 (0.0)	10 (1.4)	9.63 (.09)	-.12***	19 (5.4)	6 (1.8)	25 (3.6)	6.33 (.05)	-.10*
	없음	321 (90.7)	329 (97.9)	650 (94.2)			344 (97.2)	336 (100)	680 (98.6)			335 (94.6)	330 (98.2)	665 (96.4)		
가자림/헛소문 퍼뜨리기	있음	15 (4.2)	6 (1.8)	21 (3.0)	3.51 (.06)	-.06	1 (0.3)	0 (0.0)	1 (0.1)	0.95 (.53)	-.04	22 (6.2)	24 (7.1)	46 (6.7)	0.24 (.65)	.02
	없음	344 (95.8)	334 (98.2)	669 (97.1)			352 (95.7)	356 (100)	689 (95.9)			332 (95.8)	312 (92.9)	644 (95.3)		
금품갈취	있음	10 (2.8)	2 (0.6)	12 (1.7)	5.02 (.03)	-.09*	1 (0.3)	1 (0.3)	2 (0.3)	0.00 (.00)	.00	27 (7.6)	4 (1.2)	31 (4.5)	16.64 (.00)	-.16***
	없음	344 (97.2)	334 (99.4)	678 (98.3)			353 (97.7)	335 (99.7)	688 (96.7)			327 (92.4)	332 (98.8)	659 (95.5)		
강제삼부름/ 강요/협박	있음	7 (2.0)	5 (1.5)	12 (1.7)	.24 (.62)	-.02	2 (0.6)	0 (0.0)	2 (0.3)	1.90 (.59)	-.05	8 (2.3)	4 (1.2)	12 (1.7)	1.15 (.28)	-.04
	없음	347 (98.0)	331 (98.5)	678 (98.3)			352 (94.9)	356 (100)	688 (95.7)			346 (97.7)	332 (98.8)	678 (98.3)		
신체폭력, 외도에 대한 욕조롱	있음	64 (18.1)	28 (8.3)	92 (13.3)	14.2 (.00)	-.14***	20 (5.6)	3 (0.9)	23 (3.3)	12.10 (.00)	-.15***	43 (12.1)	20 (6.0)	63 (9.1)	7.98 (.01)	-.11**
	없음	290 (81.9)	308 (91.7)	598 (86.7)			334 (94.4)	333 (94.1)	667 (96.7)			311 (87.9)	316 (94.0)	627 (90.9)		
성적인 욕망/문짓	있음	41 (11.6)	2 (0.6)	43 (6.2)	35.61 (.00)	-.25***	10 (2.8)	0 (0.0)	10 (1.4)	9.65 (.09)	-.12***	22 (6.2)	4 (1.2)	26 (3.8)	12.00 (.00)	-.13***
	없음	313 (88.4)	334 (99.4)	647 (95.8)			344 (97.2)	356 (100)	680 (96.6)			332 (95.8)	332 (98.8)	664 (96.2)		
사이버 (휴대문) 인팅(정화)	있음	8 (2.3)	5 (1.5)	13 (1.9)	.56 (.46)	-.05	4 (1.1)	2 (0.6)	6 (0.9)	0.57 (.69)	-.05	6 (1.7)	10 (3.0)	16 (2.3)	1.25 (.26)	.45
	없음	346 (97.7)	331 (98.5)	677 (98.1)			330 (96.9)	334 (94.4)	684 (95.1)			348 (98.3)	336 (97.0)	674 (97.7)		
전체 빈도(%)	354 (100)	336 (100)	690 (100)			354 (100)	336 (100)	690 (100)			354 (100)	336 (100)	690 (100)			

*한 셀 이상이면 그 빈도를 가지는 경우는 그 셀의 통계적 유의성 검증을 위해 Fisher's exact test를 적용하였고 그 결과 p값을 이탤릭체로 표기함. 통계적으로 유의한 차이를 표인 문항은 음영으로 표시함.

였다. 엄격한 절단점을 사용한 경우, 가해율은 0.1%(거짓말/헛소문) ~ 3.6%(욕/놀림/조롱)였다. 피해율에 대한 주요 결과를 보면(표 1의 오른쪽 참조), 관대한 절단점을 사용한 경우, 피해율은 1.7%(강제심부름/강요/협박) ~ 13.6%(욕/놀림/조롱)였으며, 엄격한 절단점을 사용한 경우 피해율은 0.6%(강제심부름/강요/협박) ~ 4.6%(욕/놀림/조롱)였다. 전반적으로, 성별이나 절단점과 상관없이 중학생들에게서 가장 빈번한 유형의 괴롭힘 가해 및 피해는 언어적 괴롭힘(즉, 욕/놀림/조롱)임을 알 수 있다.

성별에 따른 가해율의 차이 및 피해율의 차이를 분석하였을 때, 주로 관계적 괴롭힘(즉, 무시/따돌림, 거짓말/헛소문)이나 상대적으로 경험빈도가 낮은 형태의 괴롭힘(즉, 사이버, 강제심부름/강요/협박)에서는 통계적으로 유의한 성차가 나타나지 않는 반면, 직접적 괴롭힘(즉, 언어, 신체, 성, 금품갈취)에서는 남학생들이 유의하게 높은 가해율과 피해율을 보였다. 파이계수를 검토할 때, 특히 언어적 괴롭힘과 관련되는 한 문항(즉, 욕/놀림/조롱)에서 성차가 가장 크게 나타나는 것을 알 수 있다. 표 1에서 확인할 수 있듯이, 이상과 같은 성차의 양상은 어떠한 절단점을 사용하여 산출하더라도 일관되게, 그리고 가해율과 피해율에서 일관되게 나타났다. 다만, 유일한 예외는 금품갈취로, 관대한 절단점을 사용한 경우, 남학생이 여학생 보다 유의하게 높은 가해율과 피해율을 보였다.

그러나 엄격한 절단점을 사용하여 반복적이고 지속적인 금품갈취에 국한시켜 산정했을 때는 성차가 더 이상 나타나지 않았다. 이는 이러한 특정 형태의 가해 혹은 피해경험 자체가 남녀 공히 매우 드물기 때문일 것으로 추론된다.

정의제시 여부와 절단점에 따른 괴롭힘 경험률의 차이

평가조건 변인 중 정의제시 여부에 따라 가해율과 피해율이 차이가 있는지를 알아보기 위해 교차분석을 실행한 결과가 표 2에 제시되어있다. 관대한 절단점과 엄격한 절단점을 적용한 경우, 그 결과를 나누어서 제시하였다. 전반질문에 따른 가해율과 피해율, 세부질문에 따른 가해율과 피해율 순으로 주요 결과를 살펴보면 다음과 같다.

전반질문에 따른 가해율의 차이는 엄격한 절단점을 사용하여 산출하였을 경우에만, 통계적으로 유의한 차이가 있었다($\chi^2 = 4.43$, $p < .05$, $\phi = .08$). 즉 정의를 제시하지 않은 평가조건에서 ‘월 2~3회 이상’의 가해경험이 있는 것으로 보고한 비율(5.6%)이 정의를 제시한 조건(2.6%)에 비해 2배 이상 높은 것으로 나타났다(증감률=111.11%). 그러나 피해율에서는 절단점과 상관없이, 정의제시 여부에 따른 유의한 차이가 관찰되지 않았다($\chi^2 = 4.43$, $p < .05$, $\phi = .08$, 증감률=38.46%). 세부질문에 따른 가해율과 피해율의 차이를 비교해 보면, 엄격한 절단점을 사용하여 경험률을 산출한 경우, 정의를 제시한 조건 대비 정의를 제시하지 않은 조건에서 욕/놀림/조롱(언어적 괴롭힘) 가해율(5.3% 대 2.0%) 및 피해율(6.5% 대 2.9%)이 모두 유의하게 높게 보고되었다(가해율 $\chi^2 = 5.36$, $p < .05$, $\phi = .09$, 증감률=157.14%; 피해율 $\chi^2 = 5.09$, $p < .05$, $\phi = .09$, 증감률=120%). 그러나 이러한 양상의 차이는 관대한 절단점을 적용한 경우에는 피해율의 경우에만 유의하게 나타났다(17.9% 대 9.4%, $\chi^2 = 10.62$, $p < .001$, $\phi = .12$, 증감률=84.85%).

요컨대, 전반적 가해율과 욕/놀림/조롱 등

표 2. 정의역시 어부와 점단점에 따른 중학생의 경험률(기해율 및 피해율)의 차이 분석 결과

문항분류 (문항번호)	기해율						피해율							
	점단점 '1-2회만'			점단점 '3회 이상'			점단점 '1-2회만'			점단점 '3회 이상'				
	경의제시 여부 미제시 빈도 (%)	전체 빈도 (%)	χ^2 (θ)	경의제시 여부 미제시 빈도 (%)	전체 빈도 (%)	χ^2 (θ)	경의제시 여부 미제시 빈도 (%)	전체 빈도 (%)	χ^2 (θ)	경의제시 여부 미제시 빈도 (%)	전체 빈도 (%)	χ^2 (θ)		
경험여부	50 (14.3)	45 (13.2)	95 (13.8)	19 (5.6)	28 (4.1)	44.5* (.04)	37 (10.6)	41 (12.1)	78 (11.3)	0.38 (.54)	15 (3.7)	18 (5.3)	31 (4.5)	1.00 (.32)
점단점	300 (85.7)	295 (86.8)	595 (86.2)	321 (94.5)	662 (95.9)		313 (89.4)	299 (87.9)	612 (88.7)		337 (96.3)	322 (94.7)	659 (95.3)	
있음	71 (20.3)	84 (24.7)	155 (22.5)	18 (5.3)	25 (3.6)	5.36* (.02)	33 (9.4)	61 (17.9)	94 (13.6)	10.62 (.00)	10 (2.9)	22 (6.5)	32 (4.6)	5.09 (.02)
없음	279 (79.7)	256 (75.3)	535 (77.5)	322 (94.7)	665 (96.4)		317 (90.6)	279 (82.1)	596 (86.4)		340 (97.1)	318 (93.5)	658 (95.4)	
무시/미동원 (가위주걱 없음)	39 (11.1)	26 (7.6)	65 (9.4)	0 (0.0)	4 (0.6)	0.00 (.00)	21 (6.0)	24 (7.1)	45 (6.5)	0.32 (.57)	8 (2.3)	5 (1.5)	13 (1.9)	.62 (.48)
없음	311 (88.9)	314 (92.4)	625 (90.6)	338 (99.4)	686 (95.4)		348 (99.4)	338 (92.9)	686 (93.5)		342 (97.7)	355 (99.5)	677 (98.1)	
매리키, 발로 차기, 타발기, 갈금 (신체폭력)	17 (4.9)	25 (6.8)	40 (5.8)	6 (1.8)	10 (1.4)	0.47 (.54)	4 (1.1)	6 (1.8)	10 (1.4)	1.19 (.28)	10 (2.9)	15 (4.4)	9 (1.3)	0.14 (.75)
없음	333 (95.1)	317 (93.2)	650 (94.2)	334 (98.2)	680 (96.6)		346 (98.9)	325 (95.6)	665 (96.4)		346 (98.9)	335 (98.5)	681 (98.7)	
거짓말/소문 퍼뜨리기	10 (2.9)	11 (3.2)	21 (3.0)	0 (0.0)	1 (0.1)	0.97 (1.0)	1 (0.3)	0 (0.0)	46 (6.7)	0.26 (.61)	7 (2.0)	6 (1.8)	15 (1.9)	.05 (.82)
없음	340 (97.1)	329 (96.8)	669 (97)	340 (100.0)	689 (95.9)		349 (99.7)	340 (95.9)	644 (93.3)		343 (95.7)	334 (98.2)	677 (98.1)	
금품갈취	6 (1.7)	6 (1.8)	12 (1.7)	0 (0.0)	2 (0.3)	1.95 (.50)	2 (0.6)	0 (0.0)	31 (4.4)	0.00 (.92)	3 (0.9)	2 (0.6)	5 (0.7)	-.06 (.68)
없음	344 (98.3)	334 (98.2)	678 (98.3)	340 (100)	688 (95.7)		348 (99.4)	340 (95.7)	659 (95.5)		347 (99.1)	338 (99.4)	685 (99.3)	
장제신분/ 강요/협박	5 (1.4)	7 (2.1)	12 (1.7)	0 (0.0)	2 (0.3)	1.95 (.50)	2 (0.6)	0 (0.0)	12 (1.7)	1.48 (.22)	2 (0.6)	2 (0.6)	4 (0.6)	.00 (.98)
없음	345 (98.6)	333 (97.9)	678 (98.3)	340 (100)	688 (95.7)		348 (99.4)	340 (95.7)	678 (98.3)		348 (99.4)	338 (99.4)	686 (99.4)	
신체폭력 외도에 대한 착오응	45 (12.3)	49 (14.4)	92 (13.3)	12 (3.5)	25 (3.6)	0.08 (.78)	11 (3.1)	12 (3.5)	65 (9.5)	1.72 (.19)	9 (2.6)	15 (4.4)	24 (3.5)	1.75 (.19)
없음	307 (87.7)	291 (85.6)	598 (86.7)	328 (96.5)	667 (96.7)		323 (92.3)	304 (89.4)	627 (90.9)		341 (97.4)	325 (95.6)	666 (96.3)	
성적인 속임/몰취	19 (5.4)	24 (7.1)	43 (6.2)	6 (1.8)	10 (1.4)	0.48 (.69)	4 (1.1)	6 (1.8)	26 (3.8)	2.81 (.62)	5 (1.4)	7 (2.1)	12 (1.7)	.40 (.53)
없음	331 (95.6)	316 (92.9)	647 (93.8)	334 (98.2)	680 (95.6)		346 (98.9)	334 (95.0)	664 (96.2)		345 (98.6)	335 (97.9)	678 (98.3)	
사이버 (휴대폰/ 인터넷/전화)	8 (2.3)	5 (1.5)	13 (1.9)	0 (0.0)	6 (0.9)	0.62 (.69)	4 (1.1)	2 (0.6)	16 (2.3)	0.32 (.62)	1 (0.3)	3 (0.9)	4 (0.6)	1.06 (.09)
없음	342 (97.7)	335 (98.5)	677 (98.1)	338 (99.4)	684 (95.1)		346 (98.9)	338 (95.1)	674 (97.7)		349 (99.7)	337 (99.1)	686 (99.4)	
전체 빈도(%)	350 (100)	340 (100)	690 (100)	340 (100)	690 (100)		350 (100)	340 (100)	690 (100)		340 (100)	340 (100)	690 (100)	

주: 빈도 이상 5보다 작은 기해 빈도를 가지는 경우는 기해율의 통계적 유의성 검증을 위해 Fisher 정황률 검증을 적용하였고 기해율과 피해율 이델리제를 표기함. 통계적으로 유의한 차이를 표기한 문항은 음영으로 표시함.

언어적 괴롭힘 가해율 및 피해율은 괴롭힘에 대한 정의가 제시되지 않는 조건에서 유의하게 높게 나타남을 알 수 있으며, 정의미제시 조건에서 경험률이 높게 나타나는 이러한 양상은, 특히 엄격한 절단점을 적용할 때(즉, 월 2~3회 이상의 보다 반복/지속적인 괴롭힘 경험을 측정할 때) 보다 크고 일관되게 나타나는 것으로 이해할 수 있다.

익명실시 여부와 절단점에 따른 괴롭힘 경험률의 차이

두 번째 평가조건 변인인 익명실시 여부에 따른 가해율과 피해율의 차이와 교차분석 결과는 표 3에 제시되어있다. 마찬가지로 관대한 절단점과 엄격한 절단점을 적용한 경우를 나누어서 경험률을 제시하였다.

표 3에서 보듯이, 모든 가해 및 피해 전반 질문 및 세부질문, 절단점을 통틀어, 익명실시 여부에 따른 통계적으로 유의한 차이를 보인 것은 관대한 절단점으로 사용하여 산출한 가해율 중 ‘거짓말/헛소문 퍼뜨리기’ 뿐이었다. 즉, 지난 2~3개월 동안 1~2회 이상 거짓말/헛소문으로 다른 사람을 가해한 경험에 대해서 기명실시 조건보다는 익명실시 조건에서 유의하게 높게 보고되었다($\chi^2 = 3.93, p < .048, \phi = .08$, 증감률=150%). 그 외의 모든 다른 가해 및 피해문항에서는 통계적으로 유의한 차이가 나타나지 않았다. 한편, 본 연구에서는 OBQ의 가해 및 피해 문항을 실시한 이후에, 익명조건과 기명조건 간에 솔직히 응답한 정도나 비밀보장을 확신하는 정도 등 심리적 경험의 차이가 유의한 차이가 있는지를 탐색해 보기 위해 사후질문을 실시하였고, 집단 간 평균차이를 비교해 보았다.

독립표본 t검증 결과, 솔직히 응답한 정도에 있어서 두 평가조건 간에 유의한 평균 차이가 없었으며[익명집단 $M(SD)=1.33(0.63)$, 기명집단 $M(SD)=1.37(0.60)$, $t=-.93, p>.05$], 응답한 내용이 다른 친구들에게 알려질 것이라고 생각하는 정도에 있어서도 두 평가조건 간 유의한 차이가 없었다[익명집단 $M(SD)=2.46(1.05)$, 기명집단 $M(SD)=2.48(1.02)$, $t=-1.85, p>.05$].

질문형태와 절단점에 따른 괴롭힘 경험률의 차이 및 분류 정확도

세 번째 평가조건 변인인 질문형태에 따라 괴롭힘 가해율 및 피해율이 차이가 있는지, 그리고 분류 정확도를 나타내는 일련의 지표는 어떠한지 살펴보기 위해, 교차분석을 실시하고, 증감률, 일치율, 이분변수 간 상관계수(ϕ 계수), 민감도, 특이도를 산출하였으며, 그 결과는 표 4에 제시되었다.

전반질문을 통해 산출된 가해율을 보면, 관대한 절단점과 엄격한 절단점 순으로 각각 13.8%, 4.1%인 반면, 세부질문을 통해 산출된 가해율은 34.2%, 6.8%로 나타났다. 전반질문을 통해 산출된 피해율의 경우, 각각 11.3%, 4.5%인 반면, 세부질문을 통해 산출된 피해율은 각각 22.9%, 7.8%로 나타났다. 질문형태에 따른 가해율의 차이 및 피해율의 차이는 통계적으로 유의하였다($\chi^2 = 116.47 \sim \chi^2 = 330.65$, 모든 $p < .01$). 질문형태에 따른 증감률을 계산해 보면, 세부질문을 통해 가해율을 산출할 경우, 전반질문 대비 67.86% ~ 148.42%, 피해율의 경우 74.19% ~ 102.56% 더 높게 나타났다. 질문형태에 따른 경험유무의 일치율은 76.09% ~ 96.09%로 엄격한 절단점을 사용하여 괴롭힘 경험률을 산정할 때, 일치율이 더

표 3. 익명성시 여부와 집단점에 따른 중학생의 경험불가해움 및 피해움의 차이 분석 결과

문항분류	가해움				피해움				
	집단점 "1-2회만"		집단점 "2-3회 이상"		집단점 "1-2회만"		집단점 "2-3회 이상"		
정답여부	익명성시 여부 인명 인도 인도 (%)								
전반질문	있음 (13.6)	47 (14.0)	48 (13.8)	95 (11.3)	0.02 (.88)	3.78 (.86)	3.78 (.86)	0.07 (.79)	0.01 (.87)
육성립(조용)	있음 (22.8)	79 (22.1)	76 (22.5)	155 (22.5)	0.05 (.81)	0.39 (.53)	0.39 (.53)	0.06 (.80)	0.12 (.73)
무시(따돌림 (기회주지 포함)	있음 (9.8)	34 (9.0)	31 (9.4)	65 (9.4)	0.13 (.71)	0.00 (.00)	0.00 (.00)	0.05 (.46)	0.32 (.40)
매리키 받본 차기, 머리기, 김금 (신체폭력)	있음 (6.4)	22 (6.4)	18 (5.2)	40 (5.8)	0.40 (.53)	1.60 (.34)	1.60 (.34)	1.01 (.32)	0.99 (.51)
가치만헛소문 퍼뜨리기	있음 (4.3)	15 (4.3)	6 (1.7)	21 (3.0)	3.93 (.05)	0.99 (1.00)	0.99 (1.00)	0.08 (.78)	0.69 (.40)
금품갈취	있음 (2.3)	8 (2.3)	4 (1.2)	12 (1.7)	1.33 (.25)	1.99 (.50)	1.99 (.50)	0.81 (.37)	5.01 (.60)
강제성부름/ 강요/점박	있음 (1.7)	6 (1.7)	6 (1.7)	12 (1.7)	0.00 (.99)	1.99 (.50)	1.99 (.50)	3.02 (.08)	0.99 (.62)
신체특징, 외도에 대한 욕/조롱	있음 (14.2)	49 (14.2)	43 (12.5)	92 (13.3)	0.42 (.52)	1.10 (.30)	1.10 (.30)	0.14 (.71)	0.67 (.41)
상적인 욕/말/몸짓	있음 (6.6)	23 (6.6)	20 (5.8)	43 (6.2)	0.21 (.65)	0.39 (.75)	0.39 (.75)	0.17 (.68)	1.38 (.24)
사이비 (유대인/ 인턴(인전화)	있음 (2.6)	9 (2.6)	4 (1.2)	13 (1.9)	1.95 (.17)	2.67 (.22)	2.67 (.22)	0.24 (.62)	0.00 (1.00)
전체 인도(%)		346 (100)	344 (100)	690 (100)		346 (100)	690 (100)		346 (100)

주: 빈 셀 이상 5보다 적은 기대 인도를 가지는 경우는 교량의 통계적 유의성 검증을 위해 Fisher 정확률검정을 적용하였고, Z값과 p값을 이탤릭체로 표기함. 통계적으로 유의한 차이를 보인 문항은 음영으로 표시함.

표 4. 질문형태와 절단점에 따른 중학생의 경험률(가해율 및 피해율)의 차이 및 분류 정확도

		가해 혹은 피해경험 절단점											
		절단점 “1~2회만”						절단점 “월 2~3회 이상”					
		전반질문/가해			전반질문/피해			전반질문/가해			전반질문/피해		
경 험 률 세 부 차 이 문	경험여부 세부\전반	있음 빈도(%)	없음 빈도(%)	전체 빈도(%)	있음 빈도(%)	없음 빈도(%)	전체 빈도(%)	있음 빈도(%)	없음 빈도(%)	전체 빈도(%)	있음 빈도(%)	없음 빈도(%)	전체 빈도(%)
	있음	83 (12.0)	153 (22.2)	236 (34.2)	75 (10.9)	83 (12.0)	158 (22.9)	16 (2.3)	31 (4.5)	47 (6.8)	29 (4.2)	25 (3.6)	54 (7.8)
	없음	12 (1.7)	442 (64.1)	454 (65.8)	3 (0.4)	529 (76.7)	532 (77.1)	12 (1.7)	631 (91.5)	643 (93.2)	2 (0.3)	634 (91.9)	636 (92.2)
	전체	95 (13.8)	595 (86.3)	690 (100)	78 (11.3)	612 (88.7)	690 (100)	28 (4.1)	662 (96.0)	690 (100)	31 (4.5)	659 (95.5)	690 (100)
χ^2 (p)		138.37(.00)			267.30(.00)			116.47(.01)			330.65(.00)		
분 류 정 확 도	증감률(%)	148.42			102.56			67.86			74.19		
	일치율(%)	76.09			87.54			93.77			96.09		
	파이계수	.45***			.62***			.41***			.69***		
	민감도(%)	35.16			47.47			34.04			53.70		
	특이도(%)	97.36			99.44			98.13			99.68		

주. 반복측정 이분변수 간 독립성 검증(χ^2)은 이항분포를 사용한 McNemar 정확확률 검증을 적용한 결과임

높았다(가해 76.09% 대 93.77%; 피해 87.54% 대 96.09%).

또한 절단점이 동일한 경우, 가해보다는 피해경험의 일치율이 더 높게 나타났다. 더욱이 이분변수 간 상관(ϕ 계수)의 크기는 절단점에 따라 큰 차이를 보이지 않았고, 가해경험의 경우 .41~.45로 대략 낮은 크기의 상관을, 피해경험의 경우 .62~.69의 중간 크기의 상관을 보였다. 끝으로, 적용된 절단점 수준과 상관없이 전반질문의 민감도는 높지 않은 반면 (34.04~53.70), 특이도는 이에 비해 일관되게 높은 것으로 나타났다(97.36~99.68). 즉, 세부 질문에 의한 괴롭힘 경험보고가 참 조건(true condition)이라고 가정한다면, OBQ의 가해 및

피해 전반질문은 실제로 괴롭힘 경험을 한 학생들 중 단지 34.04%~53.7%의 학생들만을 정확히 찾아내는 반면, 실제로 괴롭힘을 경험하지 않은 학생들 중 97.36~99.69%의 학생들을 정확히 찾아낸다고 볼 수 있다.

논 의

본 연구는 괴롭힘 실태조사에서 흔히 사용하는 자기보고식 평가방법으로 괴롭힘 경험률(즉, 가해율과 피해율)을 측정할 때 결과에 영향을 미칠 수 있는 편향요인을 탐색하는 것을 목적으로 하였다. 구체적으로 본 연구에서는

평가조건 변인들에 초점을 두어, 정의제시 여부 및 익명실시 여부에 따라 괴롭힘 경험률이 다르게 나타나는지, 또한 괴롭힘을 전반적으로 묻는 질문과 세부적으로 묻는 질문을 사용할 때 경험률에서 어떤 차이가 나타나는지를 살펴보았다. 또한 괴롭힘 경험률 평가에서 연구에 따라 적절한 절단점 수준이 다양하여 쟁점이 되어온 바, 본 연구는 서로 다른 절단점(‘관대한’ 대 ‘엄격한’ 절단점)을 적용하여 경험률을 산출한 후 그 결과를 비교해 보았다. 아울러 성별이 괴롭힘 경험의 자기보고에서 반응편향의 한 요인으로 작용할 수 있다는 증거(Rutter et al., 2003)에 기초하여, 성별에 따른 차이도 분석하였다. 본 연구는 총 690명의 남녀 중학생을 대상으로 하였으며, 이들을 설문지 지시문을 통해 실험적으로 조작된 네 가지 평가조건에 무선배정한 후, OBQ의 전반질문과 세부질문에 대한 자기보고에 기초하여 경험률을 산출하고 비교, 분석하였다. 본 연구의 주요 결과를 차례로 제시하고 논의하면 다음과 같다.

첫째, 본 연구에서 산정한 가해율과 피해율은 성별뿐 아니라 절단점 수준, 구체적인 괴롭힘 유형에 따라 다양하였다. 전반질문을 통해 측정하고, 관대한 절단점을 사용하여 괴롭힘 경험의 유무를 평가하였을 때, 가해율은 13.8%, 피해율은 11.3%였으며 엄격한 절단점을 적용한 경우는 각각 4.1% 및 4.5%로 나타났다. 성별에 따라서도 유의한 차이가 있었다. 특히 성별에 따른 경험률의 차이는 구체적인 괴롭힘 유형에 따라 달라지는 양상을 보였는데, 주로 관계적 괴롭힘(따돌림, 헛소문 등)이나 비교적 발생 빈도가 낮은 형태의 괴롭힘(강제심부름, 사이버 등)에서는 유의한 성차를 보이지 않는 반면, 직접적 형태의 괴롭힘(즉,

언어, 신체, 금품갈취, 성적인 욕 등), 특히 언어적 괴롭힘에서는 여학생에 비해 남학생의 경험률이 현저하게 높은 것으로 나타났다.

이러한 성차의 양상은 OBQ의 전반질문을 사용하여 괴롭힘 경험률을 집계한 Solberg와 Olweus(2003)의 결과와 일치한다. 또한 자기보고식의 전반질문이나 세부질문을 통해 괴롭힘 경험률이나 경험정도를 측정한 연구들에서 남학생의 가해나 피해 빈도 혹은 정도가 유의하게 높음을 보고한 국내외 선행연구들(이종원, 윤상연, 김혜진, 허태균, 2014, Camodeca & Goossens, 2002; Olweus, 1993; Sainio, Veenstra, Huitsing, & Samivalli, 2011; Salmivalli, Voeten, & Poskiparta, 2011)과도 일치하는 결과이기도 하다. 더욱이 직접적 괴롭힘에서는 남학생의 경험률이 대체로 더 높지만, 관계적 괴롭힘에서는 성차가 나타나지 않거나 여학생의 관계적 괴롭힘이나 폭력이 자기보고식 측정치에서 남학생과 유사한 수준이거나 더 높은 것으로 보고한 선행연구들(조윤오, 2013; 이은주, 2001; 이경아, 이동형, 김원희, 2015; Crick & Grotpeter, 1995)과도 전반적으로 합치되는 결과이다.

절단점을 어떻게 정하느냐에 따라 괴롭힘 경험률이 무려 2~3배 이상 차이가 있음을 보여주는 본 연구의 결과는, 단회적 혹은 산발적 수준의 가해 및 피해 경험을 반복성/지속성이 뚜렷한 ‘괴롭힘’에 포함시킬지의 여부에 따라 경험률이 크게 달라짐을 보여준다. 이는 괴롭힘을 조작적으로 정확히 정의한 후 경험률을 측정하는 것이, 괴롭힘 경험률의 고저나 증감을 논하기 위해서는 필수적임을 시사한다. 이는 또한 절단점의 차이가 괴롭힘 경험률 측정치의 신뢰도, 타당도 및 연구 간 비교가능성을 저해하는 자명한 요인이라는 지적(이동

형, 2014a; Greif & Furlong, 2006; Swearer et al., 2010)을 실증적으로 확인해 주는 결과이기도 하다.

또한 전반질문으로 측정할 경우 절단점 수준과 상관없이 가해 및 피해율은 남학생이 여학생보다 일관되게 높게 나타났지만, 구체적으로 경험한 괴롭힘의 유형별로 나누어서 질문할 경우, 주로 직접적 형태의 괴롭힘에서만 성차가 나타난 점은, 괴롭힘 경험률을 측정할 때, 구체적 유형별로 나누어 묻는 세부질문의 활용이 측정학적 이점은 차치하고라도 임상적으로 유용함을 시사한다. 만일 특정 성이 특정 형태의 괴롭힘에 취약하다면, 이러한 성차를 드러낼 수 있는 세부질문의 활용이 괴롭힘 경험에 대한 보다 정교하고 풍부한 정보를 제공할 것으로 기대할 수 있기 때문이다.

한편, 성차를 해석함에 있어서 성별에 따라 다르게 나타난 경험률은 실제적 성차를 반영할 수도 있지만, 어느 정도의 범위 내에서 성별에 따른 반응편향을 반영할 가능성도 배제할 수 없음을 기억할 필요가 있다. 본 연구에서 성차가 비교적 일관되고, 크게 나타나는 것으로 보아, 관찰된 성차가 전적으로 반응편향에 기인했을 가능성은 낮아 보인다. 그러나 반응편향에 의해 발생하는 변량의 크기는 경험률 평가 시 자기보고에만 의존하기 보다는 다른 평가방법(예를 들면, 면접, 관찰, 교사나 또래보고 등)을 통해 보완하거나 확인하는 다면적 평가 절차를 통해 추정될 수 있는 바, 성차의 변량 중 반응편향에 의해 설명되는 정도는 후속 연구를 통해 살펴볼 문제이다.

둘째, 정의제시 여부에 따른 경험률의 차이를 분석한 결과, 정의제시 조건 대비 정의미제시 조건에서 전반적 가해율 및 언어적 괴롭힘의 가해율 및 피해율이 85%~157% 더 높게

나타났다. 특히 엄격한 절단점을 적용한 경우, 두 조건 간 경험률의 차이가 보다 크고 일관되게 나타나는 경향이 있었다. 본 연구에서 두 조건 간의 차이는 모든 가해 및 피해 문항에서 일관되게 나타나지는 않았지만, 경험률 파악을 위한 전형적인 실태조사에 비해 본 연구의 표본이 작음에도 불구하고, 통계적으로 유의한 차이를 보였고, 엄격한 절단점을 적용한 경우 보다 크고 일관되게 나타났다는 점은 주목할 만하다.

Huang과 Cornell(2015, 2016)은 편향요인에 따른 괴롭힘 경험률의 차이를 해석하면서 증감률을 전형적인 개입프로그램의 효과의 크기와 비교하여 해석한 바 있다. 동일하게, 본 연구에서 나타난 85%~157%의 증감률은 학교기반의 괴롭힘 개입 프로그램의 효과성에 대한 메타분석에서 전형적인 프로그램이 괴롭힘 피해율을 평균 17%~23% 감소시킨다는 연구결과(Tofi & Farrington, 2011)에 비추어 해석해볼 수 있는데, 본 연구에서 나타난 증감률은 정의미제시가 평균적인 괴롭힘 예방 프로그램을 실시하여 얻을 수 있는 경험률 증가의 크기보다 훨씬 큰 폭의 차이를 초래할 수 있음을 보여주는 바, 매우 크고, 의미 있는 차이로 볼 수 있다. 뿐만 아니라, 괴롭힘 경험을 묻는 질문의 순서를 달리하여 괴롭힘 경험률을 평가할 때, 이러한 편향요인이 29%~76%의 피해 증감률을 보이는 것으로 보고한 Huang과 Cornell(2015)의 연구와 비교하더라도, 본 연구에서 나타난 증감률은 비교적 큰 차이로 해석할 수 있다.

아울러 괴롭힘 정의제시의 효과가 엄격한 절단점을 적용한 경우, 전반적 가해 보고율과 언어적 괴롭힘(욕, 놀림, 조롱) 가해율 및 피해율이 유의하게 높은 점은 선행연구에서 아직

보고되지 않은 흥미로운 결과로, 지속적, 반복적이며 임상적으로 보다 의미 있는 괴롭힘 경험에 국한할 때, 정의미제시가 보다 크고 일관된 방식으로 경험률 결과를 편향(과잉보고)시킬 가능성을 시사한다.

한편, 전반질문에서는 가해율의 경우에만 유의한 차이를 보인 점 또한 흥미로운 결과로 적어도 2가지 설명이 가능할 듯하다. 하나는 정의제시 조건에 비해 미제시 조건의 피해율이 높았지만 통계적 유의수준에 도달하지 못한 것은 관찰빈도가 낮은 것에 비해 표본크기가 작았기 때문일 수 있다. 또 하나의 가능한 설명은 본 연구에서 정의제시가 가해문항(총 10문항)을 제시하기 직전에 한 번만 제시되었고, 피해문항(총 10문항)을 제시하기 전에는 다시 제시되지 않았다는 점과 관련될 수도 있다. 즉 피해문항을 답변할 때는 괴롭힘에 대한 정의에 충분히 유념하여 답변하지 않았을 가능성도 있다. 이는 질문제시 순서가 괴롭힘 경험을 보고에 유의하게 영향을 미치는 것을 보여준 최근의 실증 연구결과들(Huang & Cornell, 2015, 2016)에 비추어 볼 때 가능성이 있는 설명으로, 정의제시의 효과 자체를 평가하려면 제시 순서를 무선화 혹은 역규형화하거나, 정의를 가해문항 제시 전 뿐 아니라 피해문항 제시 전에도 제시하는 실험적 절차를 적용한 후 정의제시의 효과를 검증하는 후속연구를 통해 가능한 일이다.

정의제시 조건에서 괴롭힘 경험률이 유의하게 낮게 나타난 이러한 결과는 Vaillancourt 등(2010)과 Kert 등(2009)의 연구 결과와 전반적으로 일치하는 반면, 정의제시 조건과 비제시 조건 간에 유의한 차이를 발견하지 못한 Huang와 Cornell(2015)의 결과와는 차이를 보인다. 이러한 불일치는 Huang과 Cornell(2015)의

연구에서는 Olweus의 괴롭힘 정의를 본 연구에서 보다 간단한 형태로 줄여서 제시하였을 뿐만 아니라 제시된 정의를 응답자들이 제대로 읽었는지 점검하지 않은 반면, 본 연구에서는 정의에 부합되는 경험을 보고하도록 돕는 후속 조치가 있었기 때문일 것으로 보인다. 따라서 선행연구들(Kert et al., 2009; Vaillancourt et al., 2010)과 본 연구의 결과는 괴롭힘의 정의를 발달적으로 적절한 형태로 제시하고 이를 응답자가 유념하여 읽도록 안내하였을 때, 적어도 전반적 가해율과 특정한 형태의 괴롭힘 가해율 및 피해율 평가에서 편향가능성을 줄일 수 있음을 시사하는 것으로 해석가능하다.

셋째, 익명실시 여부와 절단점에 따른 괴롭힘 경험률을 분석한 결과, 관대한 절단점을 적용하여 산출한 가해 문항 하나를 제외하고는, 모든 다른 가해 및 피해 문항에서 익명실시 여부에 따른 유의한 차이를 보이지 않았다. ‘거짓말/헛소문 퍼뜨리기’를 묻는 특정 가해경험에 대해서는 익명조건에 배정된 학생들이 보다 많이 가해경험을 보고하기는 하였으나, 이는 일회적, 산발적 수준의 가해경험이 함께 포함된 경우이고, 보다 지속적이고 반복적인 수준의 괴롭힘 가해경험에 국한하여 측정할 때는 더 이상 유의하지 않았다. 더욱이 솔직히 응답한 정도나 비밀보장을 확신하는 정도를 묻는 사후질문에서도 익명실시 조건과 기명실시 조건 간에는 유의한 평균차이가 없었다. 따라서 익명실시 조건과 비밀보장을 약속한 기명조건 간에 괴롭힘 경험률이 대체로 차이를 보이지 않으며, 특히 엄격한 절단점을 적용한 경우에는 이러한 양상이 더욱 일관되게 나타난다고 볼 수 있다.

이러한 결과는 기명 대 익명실시 조건 간에

괴롭힘 경험의 자기보고에서 유의한 차이가 없거나, 실제적으로 의미 없는, 미미한 차이가 있는 것으로 보고한 국내외의 선행연구들(최문영, 이동형, 2015; Chan et al., 2005)과 일치하는 결과이며, 괴롭힘 이외의 다른 민감한 행동의 경험률 조사에서도 비밀이 보장된다면 기명실시가 과소보고를 유도하지는 않음을 보여준 O'Malley 등(2000)의 연구결과와도 같은 맥락에서 해석될 수 있다. 또한, 두 조건 간에 괴롭힘 경험률에서 차이가 없을 뿐 아니라 함께 사후질문에서도 솔직히 보고한 정도나 실제로 비밀보장을 확신하는 정도가 다르게 나타나지 않은 점은, 괴롭힘 경험률 파악을 위한 조사에서 학생들의 솔직한 응답을 유도하기 위해 익명으로 실시하는 절차를 재고할 필요성이 있음을 보여준다.

넷째로, 질문형태와 절단점에 따른 괴롭힘 경험률을 비교한 결과, 질문형태에 따른 가해를 및 피해를 모두 유의한 차이가 있었으며, 세부질문을 사용하여 경험률을 산정한 경우 전반질문 대비 최소 68% ~ 최고 148%가 높게 나타났다. 또한 질문형태 간 괴롭힘 경험의 상관(ρ 계수)은 절단점 수준에 따라 큰 차이 없이, 가해경험의 경우 대체로 낮은 정도(.41~.45), 피해경험의 경우 중간 정도(.62~.69)의 상관을 보였다. 또한 분류 정확도의 지표로서 민감도와 특이도를 계산하였을 때, 절단점 수준에 따른 큰 차이 없이 전반질문의 민감도는 높지 않은 반면(34%~54%), 특이도(97%~99%)는 일관되게 매우 높게 나타났다.

이러한 결과는 학생들이 가해경험보다는 피해경험을 보고할 때, 전반질문과 세부질문 간에 보다 일치하는 방식으로 응답하는 경향이 있으며, 질문형태에 따라 대체로 일관되게 보고하는 경향은 있으나, 그 일관성의 정도가

매우 높은 수준은 아님을 보여준다. 더욱이 민감도와 특이도 지표는 세부질문에 따른 반응이 참이라고 가정할 때, 전반질문은 괴롭힘을 실제로 경험한 학생들의 절반 정도나 그 이하만을 정확히 찾아내는 반면, 실제로 경험하지 않은 학생들은 거의 대부분 정확히 분류됨을 보여준다.

본 연구에서 질문형태 간 괴롭힘 경험 측정치 간의 상관인 .41~.69로 나타난 점은 Solberg와 Olweus(2003)의 연구에서 이분변수 간의 상관인 가해의 경우 .77, 피해의 경우 .79 나타난 것과 비교할 때 다소 낮으나, 두 형태의 질문에 대해 서로 일치하지 않게 반응하는 응답자가 상당 수 존재함을 보여준다는 점에서 일관된 결과로 해석할 수 있다. 더욱이 민감도와 특이도 측면을 보았을 때, 전반질문은 실제로 괴롭힘 경험이 없는 학생들을 쉽게 가려내는데 효과적이지만, 실제 괴롭힘 경험이 있는 학생들을 보다 민감하게 탐지하는 목적으로는 정확도가 높지 않음을 보여주며, 이러한 결과는 Vaillancourt 등(2010)이 보고한 양상과 일치하는 결과이기도 하다. 따라서 질문형태 간의 차이에 대한 선행 연구결과와 함께 본 연구의 결과는 괴롭힘 경험률 평가에서 전반질문과 세부질문을 병행하는 방안이 분류 정확도를 높이는 차원에서 보완적 역할을 할 수 있음을 시사한다.

이상으로 논의한 본 연구의 주요 결과는 괴롭힘이나 학교폭력 평가관련 연구자나 실무자에게 몇몇 중요한 실제적 시사점을 제공한다. 이를 차례로 제시하고 논의하면 다음과 같다. 첫째, 괴롭힘 경험률은 성별은 물론이고, 절단점 수준, 구체적인 괴롭힘 유형에 따라 다양한 차이를 보이므로, 괴롭힘 경험률 연구자나 실무자는 경험률 조사 시 이러한 변인들을 분

석 변인으로 포함시키거나 통제해야 하며, 이를 통해 괴롭힘 평가의 편향가능성을 줄이도록 노력해야 한다는 점이다. 본 연구에서 절단점 수준에 따라 경험률 통계치의 차이가 무려 2~3배 이상 차이가 난 결과가 보여주듯이, 절단점은 괴롭힘 평가 결과의 비일관성을 초래하는 가장 자명하고 쉽게 통제가 가능한 변인이므로, 이를 고려하지 않는 괴롭힘 경험률 논의는 무의미할 가능성이 크다. 따라서 괴롭힘 연구자들은 각 연구나 조사에서 절단점을 명확히 정의하여 경험률을 집계해야 하며, 여러 실태조사 결과를 비교할 경우에도 절단점 수준에 유의하고, 서로 다른 실태조사 간에도 적용 절단점을 최대한 일관되게 유지하거나 통일하는 노력을 우선적으로 기울여야 할 것이다. 또한 대부분의 문항에서 뚜렷한 성차가 나타난 바, 이러한 차이는 괴롭힘 경험에서의 실제적인 성차를 반영할 수도 있지만 동시에 어느 정도의 반응편향에 기인할 가능성도 있으며, 이를 감안하여 성차를 해석할 필요도 있다.

둘째, 괴롭힘 경험률 측정치의 신뢰성, 타당성 및 정확성을 제고하기 위해서는 무엇보다도 괴롭힘의 정의를 명확히 하여 발달적으로 적절한 정의를 제시하고, 본 연구에서처럼 제시된 정의를 반드시 읽도록 안내하는 절차나 유사한 절차를 추가하는 것을 고려해야 할 것이다. 이는 본 연구결과가 정의제시 여부에 따라 경험률 측정치가, 일부 문항에서 85%~157%의 차이를 보였으며, 특히 단순한 또래 간의 산발적 폭력경험 이상의 지속성/반복성을 특징으로 하는 보다 심각한 괴롭힘 경험을 측정할 경우, 정의제시가 과잉보고를 예방할 가능성을 보여주기 때문이다. 아울러, 단순히 괴롭힘의 정의를 제시한 선행 연구

(Huang & Cornell, 2015)에서 정의제시 여부가 유의한 영향을 미치지 않은 것과 비교할 때, 제시된 정의를 학생들이 반드시 숙지하도록 유도하는 절차의 중요성 또한 유념할 필요가 있다.

셋째, 괴롭힘 경험률 평가 시, 비밀보장을 전제로 한 기명실시 절차를 적극 고려할 필요가 있다. 본 연구는 지속적/반복적 수준의 괴롭힘 경험률을 측정할 때, 비밀보장을 약속하고(즉, 친구들에게 알려지지 않음을 약속함), 기명으로 실시하더라도(즉, 학년, 반, 번호를 포함한 개인정보를 기입하게 하는 방식), 괴롭힘 경험 보고가 편향될 가능성이 거의 없고, 솔직히 응답한 정도나 비밀보장을 확신하는 정도에 있어서도 익명조건과 차이가 없음을 보여주었다. 서론에서 논의한 바와 같이 기명 실시 절차는 타당성 확인이나 평가정보의 활용 측면에서 뚜렷한 장점이 있다. 특히 괴롭힘 실태조사의 주목적이 적절한 ‘개입’으로 연결하는 것이라면, 익명실시는 괴롭힘 경험이 있는 학생에 대한 후속 개입조치를 철저히 어렵게 하는 모순적 절차임을 기억할 필요가 있다. 따라서 괴롭힘 평가 연구자나 실태조사 실무자들은, ‘학생의 발달수준에 적절하고, 학교장면의 특성을 반영한, 비밀보장을 전제로 한 기명실시 절차’를 개발하고 이러한 평가조건 하에서 괴롭힘 경험률을 평가하여 측정의 타당성을 높이고, 후속 개입과 연계하는 노력을 기울여야 할 것이다.

넷째, 괴롭힘 경험률 평가 시, 본 연구의 OBQ와 같이 괴롭힘 경험에 대한 전반질문과 함께 세부질문을 병용함으로써 얻을 수 있는 측정학적, 임상적 이점을 고려할 필요가 있다. 본 연구결과는 전반질문과 세부질문을 통한 경험보고가 서로 중복되지만 낮거나 중간정도

의 상관에 그치며, 두 형태의 질문이 임상적으로나 측정학적으로 상보적 역할을 할 수 있음을 시사하였다. 다만, 본 연구의 결과에서처럼 괴롭힘 경험률이 질문형태에 따라 적지 않은 차이를 보일 수 있으므로, 이러한 차이를 어떻게 통합하여 해석하고 활용할지에 대한 연구기반의 지침이 마련될 때 까지는, 두 질문형태에 따른 경험률 측정치를 병행하여 제시하는 방안이 바람직할 것으로 보인다(예컨대, 본 연구에서 괴롭힘 가해율 추정치는 4.1%~6.8%, 피해율 추정치는 4.5%~7.8%). 이러한 방안이 적어도 단일 측정치에 기초하여 경험률의 증감을 평가하고 논하는 것에 비해, 보다 현실적일 뿐 아니라 신뢰로울 가능성이 크기 때문이다.

본 연구의 제한점은 괴롭힘 경험률에 영향을 미치는 요인을 연구한 외국의 대표적인 연구들(예: Solberg & Olweus, 2003; Huang & Cornell, 2015, 2016)이나 국내의 학교폭력 실태조사 등의 표본에 비해 본 연구의 표본크기(690명)가 상대적으로 작다는 점이다. 특히 OBQ에서 일반적으로 권고되는 ‘월 2~3회 이상’의 괴롭힘 경험은 관찰빈도가 보다 낮아서, 본 연구의 경우 상대적으로 드문 특정 유형의 괴롭힘의 경우, 관찰빈도가 1~2 사례뿐인 경우도 있었다. 아울러 표집의 범위가 영남권의 7개교에 재학 중인 중학생에 국한된 점 또한, 표본의 대표성과 결과의 일반화가능성 면에서 한계점으로 남는다. 그러나 본 연구의 주목적이 경험률을 정확히 추정하는 것이 아니라, 잠재적인 편향요인의 효과를 평가하는 것이었으므로, 상대적으로 작은 표본을 통해서도 각 요인의 편향 가능성은 적절히 탐색할 수 있었다고 본다. 본 연구에서는 높고 낮은 수준의 두 가지 절단점을 적용한 분석 결과를 함께

제시하여 절단점에 따른 차이를 비교하였고, 셀 당 기대빈도가 5미만인 경우에는 Fisher 정확률검증을 적용하여 유의성을 검증하였다. 이후 보다 크고 대표적인 표본을 활용하여 잠재적 편향요인의 효과를 살펴보는 후속연구가 이어진다면, 본 연구에서 나타난 편향요인의 효과를 더욱 신뢰롭게 평가해 볼 수 있을 것이다.

이처럼 표본의 크기의 한계와 탐색적 연구로서 갖는 제한점에도 불구하고, 본 연구는 자기보고식 괴롭힘 평가 분야에서 생점이 돼 온 많은 잠재적 편향요인들 중 평가조건 관련 몇몇 요인들을 선정하여 무선배정에 의한 실험적 방법을 통해 괴롭힘 경험률에 미치는 구체적 영향을 실증적으로 검증하였다는 점에서 의의가 있다. 특히, 괴롭힘 경험률 평가의 편향요인에 대해서 국외의 괴롭힘 연구문헌에서 꾸준히 논의가 있었고, 최근 들어 국내의 연구문헌도 이러한 부분을 다루고 있지만, 이를 실증적으로 평가한 연구물은 국내는 물론이고 국외의 경우에도 많지 않은 상황이다. 특히 OBQ식의 경험률 측정과 관련하여 잠재적 편향요인의 영향을 실험적 방법으로 평가한 국내 실증연구는 아직 없었는데, 본 연구는 국내 중학생 표본을 대상으로 정의제시나 익명 실시 여부, 질문형태, 절단점 등 괴롭힘 평가의 잠재적 편향요인으로 거론돼 몇몇 요인들이 경험률 측정치에 미치는 영향을 실험적 방법을 통해 직접 검증함으로써 이러한 변인들의 영향을 살펴본 국외의 연구문헌들과 비교할 수 있는 실증 자료를 제시하였다는 점에서 의의가 있다.

또한 괴롭힘이나 학교폭력의 증감에 대해 사회적으로나 교육현장의 시각이 민감하지만, 이에 비해 괴롭힘이나 학교폭력의 평가와 관련

된 실증연구가 아직 미미한 국내의 실정에서, 본 연구는 괴롭힘이나 학교폭력 연구자뿐 아니라 실태조사를 개발하고 적용하며, 평가 절차를 개선하고자 하는 관련 실무자들에게도 구체적인 시사점을 제공하고 있다. 본 연구가 괴롭힘 평가 관련 후속 실증연구를 자극하고, 아울러 국내 괴롭힘이나 학교폭력의 경험을 관련 연구나 실태조사의 과학적 엄밀성을 제고하는 데 기여할 수 있기를 기대해 본다.

참고문헌

- 교육부 (2017). 2017년 1차 학교폭력 실태조사 결과 - 보도자료. <http://moe.go.kr>에서 2018, 8, 20 인출.
- 김혜원 (2013). 청소년 학교폭력: 이해, 예방, 개입을 위한 지침서. 서울: 학지사.
- 신유림 (2006). 학령기 아동의 또래괴롭힘에서 측정방법에 따른 심리사회적 부적응 차이: 자기보고법과 또래 지명법의 비교. *한국가정관리학회지*, 24(4), 141-149.
- 이규미, 지승희, 오인수, 송미경, 장재홍, 정제영, 조용선, 이정윤, 유형근, 이은경, 고경희, 오혜영, 이유미, 김승혜, 최희영 (2014). 학교폭력 예방의 이론과 실제. 서울: 학지사.
- 이경아, 이동형, 김원희 (2015). 남녀 초등학생의 공감 및 지각된 학급풍토와 괴롭힘 참여역할의 관계. *교육문제연구*, 28(3), 21-47.
- 이동형 (2014a). 학교 괴롭힘의 개념화와 자기보고식 평가: 현 쟁점 및 개선 방안 고찰을 중심으로. *한국심리학회지: 일반*, 33(4), 737-763.
- 이동형 (2014b). 학교폭력의 평가와 진단. 이화여자대학교 학교폭력예방연구소 편 “학교폭력과 괴롭힘 예방: 원인진단과 대응” (10장). 서울: 학지사.
- 이종원, 윤상연, 김혜진, 허태균 (2014). 권위주의 성격과 인기도에 따른 학교괴롭힘의 참여 역할. *한국심리학회지: 학교*, 11(1), 109.
- 정태연, 김인경, 김은정 (2001). 집단 따돌림의 측정방법에 따른 행동평가의 차이: 자기평가와 동료평가의 비교. *한국심리학회지: 발달*, 14(1), 145-149.
- 조운호 (2013). 성별에 따른 학교폭력 발생실태 및 유발요인 차이 분석. *청소년복지연구*, 15(1), 155-179.
- 이은주 (2001). 공격행동의 유형 및 성별에 따른 집단 괴롭힘 가해행동과 피해아동의 또래관계 비교. *아동학회지*, 22(2), 167-180.
- 최문영, 이동형 (2015). 괴롭힘 참여역할 자기보고의 잠재적 편향요인: 개인특성 및 평가조건 변인을 중심으로. *한국심리학회지: 학교*, 12(2), 203-224.
- 청소년폭력예방재단 (2018). 2017 전국 학교폭력 실태조사 연구. (재)푸른나무 청재단.
- 최희영 (2008). 괴롭힘 상황에서의 아동의 가해 및 피해경험 측정방법 비교를 통한 심리적 특성 연구. *초등교육연구*, 21(3), 339-358.
- Baly, M. W., & Cornell, D. G. (2011). Effects of an educational video on the measurement of bullying by self-report. *Journal of School Violence*, 10, 221-238.
- Batsche, G. M. (1997). Bullying. In G. G. Bear, K. M. Minke, & A. Thomas (Eds.), *Children's*

- needs, Part II: Development, problems, and alternatives* (pp. 171-179). Bethesda, MD: National Association of School Psychologists.
- Camodeca, M., Gossens, F. A., Schuengel, C., & Meerum, T. M. (2003). Links between social information processing in middle childhood and involvement in bullying. *Aggressive Behavior, 29*, 116-127.
- Chan, H. F. J., Myron, R., & Crawshaw, M. (2005). The efficacy of non-anonymous measures of bullying. *School Psychology International, 26*, 443-458.
- Cornell, D. G. & Cole, J. C. M. (2012). Assessment of bullying. In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), *Handbook of school violence and school safety: International research and practice* (pp. 289-303). New York, NY: Routledge.
- Cornell, D. G. & Cole, J. C. M. (2012). Assessment of bullying. In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), *Handbook of school violence and school safety: International research and practice* (pp. 289-303). New York, NY: Routledge.
- Cornell, D. G., & Bandyopadhyay, S. (2010). The assesment of bullying. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 265-276). New York, NY: Routledge.
- Cornell, D. G., Sheras, P. L. & Cole, J. C. (2006). Assessment of bullying. In S. R. Jimerson & M. J. Furlong (Eds.), *Handbook of school violence and school safety* (pp.587-602). Mahwah, NJ: Erlbaum.
- Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social psychological adjustment. *Child Development, 66*, 710-722.
- Crick, N. R., & Grotpeter, J. K. (1996). Children's treatment by peers: Victims of relational and overt aggression. *Development and Psychopathology, 8*, 367-380.
- Crothers, L. M., & Levinson, E. M. (2004). Assessment of bullying: A review of methods and instruments. *Journal of Counseling & Development, 82*(4), 496-503.
- Furlong, M. J., Sharkey, J. D., Felix, E. D., Tanigawa, D., & Green, J. G. (2010). Bullying assessment: A call for increased precision of self-reporting procedures. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp.329-345). New York, NY: Routledge.
- Greif, J. L. & Furlong, M. J. (2006). The assessment of school bullying: Using theory to inform practice. *Journal of School Violence, 5*, 33-50.
- Hamby, S. L., & Finkelhor, D. (2000). The victimization of children: Recommendations for assessment and instruction development. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 829-840.
- Huang, F. L., & Cornell, D. G. (2015). The impact of definition and question order on the prevalence of bullying victimization using student self-reports. *Psychological Assessment, 27*(4), 1484-1493.
- Huang, F. L., & Cornell, D. G. (2016). Question order affects the measurement of bullying

- victimization among middle school students. *Educational and Psychological measurement*, 76(5), 724-740.
- Jia, M., & Mikami, A. (2018). Issues in the assessment of bullying: Implications for conceptualizations and future directions. *Aggression and Violent Behavior*, 41, 108-118.
- Kert, A. S., Coddling, R. S., Tyron, G. S., & Shiyko, M. (2009). Impact of the word "bully" on the reported rate of bullying behavior. *Psychology in the Schools*, 47, 193-204.
- Land, D. (2003). Teasing apart secondary students' conceptualizations of peer teasing, bullying, and sexual harassment. *School Psychology International*, 24, 147-165.
- Menesini, E., & Salmivalli, C. (2017). Bullying in schools: The state of knowledge and effective interventions. *Psychology, Health & Medicine*, 22(sup1), 240-253.
- O'Malley, P. M., Johnston, L. D., Bachman, J. G., & Schulenberg, J. E. (2000). A comparison of confidential versus anonymous survey procedures: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *Journal of Drug Issues*, 30, 35-54.
- Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Oxford, UK: Blackwell Publishers.
- Olweus, D. (2007). *The Olweus Bullying Questionnaire*. Center City, MN: Hazelden.
- Olweus, D. (2010). Understanding and researching bullying: Some critical issues. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp.9-34). New York, NY: Routledge.
- Rutter, M., Caspi, A., & Moffitt, T. E. (2003). Using sex differences in psychopathology to study causal mechanisms: Unifying issues and research strategies. *Journal of Child Psychology and Psychiatry*, 44(8), 1092-1115.
- Sainio, M., Veenstra, R., Huising, G., & Salmivalli, C. (2011). Victims and their defenders: A dyadic approach. *International journal of behavioral development*, 35(2), 144-151.
- Salmivalli, C., Voeten, M., & Poskiparta, E. (2011). Bystanders matter: Associations between reinforcing, defending, and the frequency of bullying behavior in classrooms. *Journal of Clinical Child & Adolescent Psychology*, 40(5), 668-676.
- Sharkey, J. D., Dowdy, E., Twyford, J. Twyford, J., & Furlong, M. J. (2012). An overview of measurement issues in school violence and school safety research. In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), *Handbook of school violence and school safety: International research and practice* (pp. 259-272). New York, NY: Routledge.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioural sciences*(2nd. ed). New York: McGraw-Hill.
- Smith, P. K., Cowie, H., Olafsson, R. F., & Liefoghe, A. P. D. (2002). Definitions of bullying: A comparison of terms used, and age and gender differences, in fourteen-country international comparison. *Child Development*, 73, 1119-1133.
- Solberg, M., & Olweus, D. (2003). Prevalence

- estimation of school bullying with the Olweus bully/victim questionnaire. *Aggressive Behavior*, 29, 239-268.
- Sveinsson, A. V., & Morris, R. J. (2007). Conceptual and methodological issues in assessment and intervention with school bullies. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization, and peer harassment: A handbook of prevention and intervention* (pp.9-26). New York, NY: Routledge.
- Swearer, S. M., Siebecker, A. B., Johnsen-Frerichs, L. A., & Wang, C. (2010). Assessment of bullying/victimization: The problem of comparability across studies and across methodologies. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 305-327). New York, NY: Routledge.
- Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, 7(1), 27-56.
- Vaillancourt, T., Trinh, V., McDougall, P., Duku, E., Cunningham, L., Cunningham, C., Hymel, S., & Short, K. (2010). Optimizing population screening of bullying in school-aged children. *Journal of School Violence*, 9, 233-250.
- Volk, A. A., Dane, A. V., & Marini, Z. A. (2014). What is bullying? A theoretical redefinition. *Developmental Review*, 34(4), 327-343.
- 원고접수일 : 2018. 10. 23.
수정원고접수일 : 2018. 12. 16.
최종게재결정일 : 2018. 12. 21.

Biasing Factors in Self-Report Assessment of Bullying/Victimization: Examining Variability in Involvement Rates by Testing Conditions

Donghyung Lee

Department of Education, Pusan National University

The self-report assessment has been most commonly used to estimate bullying/victimization (B/V) rates in most domestic and international prevalence studies. However, the presence of many potential biasing factors in such an assessment method, including specific operationalization/measurement strategies and testing conditions, has become an issue due to a considerable variability in reported involvement rates across studies. This study analyzed self-reported B/V involvement rates on Olweus Bullying Questionnaire (OBQ) among 690 Korean middle school students by gender and two different cut-offs (generous vs. strict cut-offs) and examined if the involvement rates were significantly varied by testing conditions such as presentation vs. omission of a precise definition of B/V, anonymous vs. non-anonymous/confidential administration, and the use of global vs. specific questions. Chi-square analyses revealed that boys displayed higher involvement rates on global measures of B/V and on items related to direct forms of B/V, with no significant gender differences on specific measures of relational B/V rates. It was also found that a global rate of bullying and specific rates of verbal B/V were 111% to 157% higher when no definition was provided. However, anonymous vs. non-anonymous administration had no significant impacts on rates of involvement, except for one item; there were also no significant differences in reported degrees of frankness and perceived confidentiality of their responses across two administration conditions. Finally, when involvement rates were assessed by using specific vs. global items, they were 68% to 148% higher with binominal correlations in low to moderate ranges. Findings also indicated that global items had a high specificity but a relatively low sensitivity. Implications of these findings were fully discussed for researchers and practitioners in the field of B/V assessment.

Key words : bullying, self-report assessment, school violence, biasing factors