

학교 괴롭힘의 개념화와 자기보고식 평가: 현 쟁점 및 개선방안 고찰을 중심으로*

이 동 형[†]

부산대학교 교육학과

본 연구는 학교에서의 괴롭힘에 대한 연구물이 국내외적으로 급증하고 있는 상황에서 괴롭힘의 개념화 및 평가 관련 주요 쟁점들을 정리하고 개선방안을 고찰하였다. 이를 위해 괴롭힘에 대한 국내외 최근 연구문헌들을 검토하여 괴롭힘의 개념화에 대한 현 쟁점을 핵심 요소, 하위 유형 및 용어 선택의 문제로 나누어 자세히 분석하고, 괴롭힘 연구에서 현재 널리 사용되고 있는 자기보고식 평가방법을 중심으로 괴롭힘의 조작화 및 측정 방법론적 쟁점들을 핵심 요소 및 하위 유형의 조작화, 단일문항 대 다문항 접근, 실시 맥락과 절차, 응답자의 반응 특성, 공유방법분산의 문제로 범주화하여 차례로 고찰하였다. 또한 이러한 고찰을 바탕으로 국내외 괴롭힘 연구자들이 유념해야 할 시사점을 도출하고 괴롭힘 평가의 향후 개선방안을 제언하고 논의하였다.

주요어 : 괴롭힘, 학교폭력, 개념화, 조작화, 자기보고식 평가

* 본 연구는 2013학년도 부산대학교 교내 학술연구비(신임교수연구정착금)에 의한 연구임.

† 교신저자: 이동형, 부산대학교 교육학과, Tel: 051-510-2638, E-mail: leepsy@pusan.ac.kr

학교폭력(school violence)이 교육현장의 안전과 학생들의 건강한 발달을 위협하는 중대한 사회적 문제로 대두됨에 따라 학교 괴롭힘(bullying in schools 혹은 school bullying)에 대한 연구가 국내에서도 활발히 진행되고 있다. 괴롭힘에 대한 학술적 연구의 기원은 1960년대 말 스웨덴의 Daniel Olweus가 시작한 일련의 종단연구에서 찾을 수 있으며, 1980년대 노르웨이에서 발생한 일련의 학생 자살사건이 괴롭힘과 관련된 것으로 나타남에 따라 전국적 연구로 확대되었고, 1990년대를 지나면서 점차 유럽 전역의 관심과 미국을 포함한 국제적 관심을 받기에 이르렀다(Sveinsson & Morris, 2007). 이후로 전 세계적으로 괴롭힘에 대한 연구문헌들이 급증하고 있는 추세에 있으며, 이러한 추세는 국내의 경우에도 뚜렷하게 나타나고 있다. 단적으로, 괴롭힘 및 관련 검색어(‘학교폭력’ 혹은 ‘따돌림’)를 키워드로 RISS 검색엔진을 사용하여 국내학술지에 발표된 논문 수를 검색해 보면, 1990년대에 399편에 불과하던 것에 비해 2000년대에는 1329편, 2011년 이후로 2014년 중반까지 최근 약 3~4년 사이에 발표된 학술지 논문 수만 이미 1260편을 상회하고 있다.

괴롭힘은 정의하고 평가하기가 비교적 쉽지 않은 개념이라는 점에 대해 많은 학자들이 동의한다(Cornell, Sheras, & Cole, 2006; Griffin & Gross, 2004). 그러나 괴롭힘에 대한 연구가 이 같이 급증하고 있는 가운데 괴롭힘에 대한 과학적 이해와 지식의 진보를 위해서는 무엇보다도 괴롭힘에 대한 명확한 개념화를 토대로 측정의 정확성을 높이는 일이 매우 중요하다. 한 현상에 대한 타당한 측정과 평가는 모든 과학적 연구에서 필수적인 요소이지만, 특히 괴롭힘과 같이 사회적으로 큰 관심을 받고 있

는 현상의 경우에는 실제적 이유에서 그 중요성이 더욱 강조될 수밖에 없다. 이는 괴롭힘 발생의 실태나 특징을 정확하게 파악하여 이러한 정보가 괴롭힘의 예방과 개입을 위한 다양한 수준에서의 노력에 효과적으로 반영되어야 하고, 또 이러한 노력들의 성과에 대한 객관적인 평가를 기반으로 지속적인 개선의 노력을 해 나가야 괴롭힘과 같은 사회병리적 문제에 효과적으로 대처할 수 있기 때문이다. 괴롭힘의 측정 및 평가 관련 주제는 괴롭힘 분야의 대표적인 최근 편저들(Jimerson, Nickerson, Mayer & Furlong, 2012; Jimerson, Swearer, & Espelage, 2000; Zins, Elias, & Maher, 2007)에서 비중 있게 다루어지고 있는 주제 중 하나로 괴롭힘에 대한 학술적, 사회적 관심이 확대되고 있는 국내의 상황에서도 반드시 짚고 넘어가는 것이 시의적절하다고 할 수 있다.

국내의 경우, 괴롭힘 관련 용어들이 명확한 개념적 구분 없이 혼용되는 경향이 있으며, 이러한 용어상의 혼란과 맞물려 괴롭힘 혹은 유사 개념의 측정도구들에서 사용되는 조작적 정의나 측정방식 또한 일관성보다는 다양성을 보이고 있는 실정이다. 이러한 상황은 외국의 경우도 크게 다르지 않아 보인다. 가령, Swearer, Siebecker, Johnsen-Frerichs와 Wang(2010)은 괴롭힘 가해/피해 유병률을 보고한 연구들을 개관하였는데, 연구에 따라서 적게는 13%에서 많게는 75%까지 큰 차이를 보인다는 사실을 지적하면서, 이처럼 유병률이 연구에 따라 큰 차이를 보이는 원인이 괴롭힘의 정의 및 측정 방법론적 문제와 상당 부분 관련되어 있다고 분석하였다. 괴롭힘 현상 자체를 밝히는 연구, 괴롭힘 관련 변인에 대한 연구, 괴롭힘 개입에 대한 실증연구는 계속 늘어나는 추

세에 있는 반면, 이러한 연구들에서 사용되고 있는 측정도구의 타당도에 대한 연구나 괴롭힘의 측정 및 평가와 관련된 주제를 다룬 국내 실증연구는 자기보고식 평가와 또래평정법 혹은 교사평정법을 비교한 소수의 연구(신유림, 2006; 정태연, 김인경, 김은정, 2001; 최지영, 2008)가 있을 뿐이다. 또한 괴롭힘의 측정 및 평가와 관련된 최근 쟁점을 체계적으로 정리하거나 소개한 개관연구도 아직 없는 실정이다. 따라서 본 연구는 괴롭힘 실태조사나 학술연구에서 가장 보편적으로 사용되고 있는 자기보고식 평가에 초점을 맞추어 괴롭힘 평가와 관련된 최근 쟁점을 주로 외국의 학술문헌을 토대로 고찰하고, 이러한 쟁점들이 국내의 괴롭힘 연구자들에게 시사하는 점은 무엇이며, 향후 개선방안은 어떠한지 논함으로써 국내 괴롭힘 연구의 과학적 엄밀성을 높이는 데 기여하고자 한다. 이를 위해 본 연구는 먼저 괴롭힘의 개념화와 관련된 현 쟁점을 검토하고 이를 토대로 괴롭힘의 자기보고식 평가와 관련된 쟁점을 분석하고 시사점과 개선방안을 제시할 것이다.

개념화의 쟁점

괴롭힘에 대한 명확한 개념정의 없이는 괴롭힘의 타당한 측정과 평가를 논하기 어렵다. 따라서 이 절에서는 괴롭힘 평가의 문제를 다루기에 앞서, 괴롭힘의 개념화와 관련된 현 쟁점을 살펴보고자 한다. 현재 괴롭힘 개념화의 쟁점은 상호 연관된 세 가지 문제로 집약해 볼 수 있다. 괴롭힘의 본질적 요소가 무엇인가 하는 점, 괴롭힘에 포함되어야 할 행동의 범위는 어떠한지 하는 점, 그리고

괴롭힘 현상을 지칭하는 적절한 용어는 무엇인가 하는 점이 그것이다.

핵심 요소

가장 보편적으로 받아들여지는 괴롭힘의 학술적 정의는 괴롭힘 연구의 초석을 놓은 Olweus(1993)의 정의다. 그에 따르면 괴롭힘은 다른 사람에게 상해를 가하거나 불편을 주고자 의도한 반복적이고 지속적인 공격행동의 한 형태로, 피해학생이 자신을 방어하기 어려운 특징이 있다(Olweus, 2010). 이러한 정의는 ‘의도성’과 ‘지속성/반복성’ 뿐 아니라 가해학생과 피해학생 간에 존재하는 ‘힘의 불균형’을 포함한 3대 핵심 요소를 강조한다. 그러나 이러한 요소들에 대해서 다소 다른 견해를 보이는 학자들도 있다. 예컨대, Ross(2003)는 공격행동이 일회적인 것이든 반복적인 것이든 피해자가 괴롭힘을 당했다는 지각이 고려되어야 한다고 주장하였으며, 유사하게 Arora(1996)도 일회적인 공격행동으로도 피해자는 외상을 경험할 수 있는 경우가 있으므로 반복성보다는 공격행동의 장기적 영향력을 핵심 요소로 보아야 한다고 주장하였다. 또한 Rigby(2004)는 Olweus(1993)의 3대 요소에 동의하지만 행위의 유해성과 부당성, 피해학생이 느끼는 중압감, 가해학생이 자신의 행위를 즐기는 점을 추가해 7가지 핵심 요소를 제안하였으며, Griffin과 Gross(2004)는 괴롭힘은 피해학생의 도발에 의해 촉발된 것이 아니라는 점, 또래들이 존재하는 상황에서 또래집단 내에서 발생한다는 점을 추가 요소로 포함시켜야 한다고 주장하였다.

이처럼 괴롭힘을 정의하는 핵심 요소를 둘러싼 쟁점은 괴롭힘 개념화의 복잡성을 보여

주지만, 학자들 간에 대체로 이견이 없는 점은 괴롭힘이 의도적인 공격행동의 한 하위 유형이며, 가해자와 피해자 사이에 신체적인 힘, 운동 능력, 또래지위나 인기도 등의 차이에서 비롯되는 힘의 불균형이 존재한다는 점, 혹은 Crick과 Dodge(1999)에 표현을 사용한다면, ‘피해자에 대한 가해자의 대인관계적 지배성’이 내포된다는 점이다. 의도성은 괴롭힘을 또래 간의 갈등이나 거친 장난 등과 구별 짓는 요소이며, 힘의 불균형은 ‘또래 공격행동’을 나타내는 보다 포괄적인 용어들(대표적으로 ‘peer aggression’ 혹은 ‘peer victimization’이라는 용어)로부터 괴롭힘의 특수성을 드러내는 핵심 요소라는 점이 이 분야의 여러 학자들에 의해 일관되게 지적되고 있다(Furlong, Sharkey, Felix, Tanigawa, & Green, 2010; Greif & Furlong, 2006; Olweus, 2010; Swearer, Siebeck, Johnsen-Frerichs, & Wang, 2010). 예컨대, Swearer 등(2010)은 힘의 불균형에 대해 피해자로 하여금 스스로를 방어할 수 없게 만드는, 괴롭힘 역동에 ‘내재된’ 요소라고 기술하였으며, Olweus(2010) 또한 괴롭힘 정의와 측정에 이 요소를 포함시키지 않을 경우 모든 형태의 또래 공격행동이 모두 괴롭힘 가해/피해경험으로 이해될 수 있을 뿐 아니라 가해자와 피해자 역할 간의 중복을 증가시켜 가해자와 피해자가 대략 동일하며 비슷한 학생들이라는 부정확한 결론에 도달할 수 있다고 경고하기도 하였다. 요컨대, 괴롭힘의 본질적인 요소로서 의도성과 힘의 불균형은 가장 폭넓게 동의되는 요소인 반면, 반복성을 포함시킬지의 여부, 그리고 그 외의 요소들이 꼭 필요한지에 대해서는 이견들이 존재하는 것으로 볼 수 있다.

하위 유형

괴롭힘을 정의하는 요소와 함께 자주 거론되는 쟁점은 괴롭힘에 포함되는 구체적인 공격적 행동의 하위 유형은 무엇인가 하는 점이다. 괴롭힘이라는 특수한 형태의 또래 공격행동을 기술하면서 이 분야의 대표적인 연구자들은 ‘직접적’ 형태의 괴롭힘(면대면의 신체적 혹은 언어적 가해행동)뿐 아니라 ‘간접적’ 형태의 괴롭힘(소문을 퍼트리거나 사회적 배제와 같은 덜 가시적인 가해행동)까지 포함시켜 왔으며, 최근에는 사이버괴롭힘도 한 하위 유형으로 포함시키고 있다(Olweus, 2007). 여컨대, Olweus(2007)는 아홉 가지의 하위 유형을 언급하고 있는데, 여기에는 험담이나 욕설 같은 언어적 괴롭힘, 때리기, 차기, 밀기, 침뱉기 같은 신체적 괴롭힘, 사회적 배제나 소외를 통한 괴롭힘, 거짓말이나 헛소문을 통한 괴롭힘, 돈이나 물건을 뺏거나 훼손하는 행위, 어떤 일을 하도록 강요하거나 협박하는 행위, 인종적 괴롭힘, 성적 괴롭힘, 사이버 괴롭힘이 포함된다.

이러한 개념화는 괴롭힘의 상위 개념인 또래 공격행동에 대한 연구자들이 외현적 공격행동과 더불어 덜 외현적 형태의 공격행동(예: 관계적 공격)을 구분하고 성차 연구 등을 통해 이러한 개념적 구분의 타당성을 입증함에 따라, 괴롭힘의 개념화에도 자연스럽게 반영된 것으로 해석된다. 그러나 괴롭힘에 포함되는 공격행동의 유형을 이렇게 넓게 보는 것에 대해 우려하는 시각도 있다. 가령, Sveinsson과 Morris(2007)는 간접적 형태의 괴롭힘은 구체적인 행동적 지표가 직접적 형태의 괴롭힘에 비해 모호할 뿐 아니라, ‘또래배척’이나 ‘관계적 공격’ 등의 다른 용어로 더 잘 기술할 수 있

는 공격행동을 굳이 괴롭힘에 포함시켜야 하는지 의문을 제기하면서, 괴롭힘 개념을 지나치게 포괄적으로 접근하는 것은 비생산적이라고 주장하였다. 또한 사이버괴롭힘의 경우 직접적 형태의 괴롭힘을 정의할 때 언급되는 3가지 핵심 요소(특히, 반복성과 힘의 불균형)에 준하여 개념을 정의하기 어려운 측면들이 있어서 괴롭힘의 하위 유형으로 포함시키는 것이 적절한지의 여부에 대해 논란이 되고 있다(Smith, 2012). 같은 맥락에서, Smith, Cowie, Oafsson과 Liefoghe(2002)는 영어의 ‘bullying’이라는 용어가 다양한 언어로 13개 언어로 번역되면서 14개국에서 사용되는 번역용어의 의미 차이를 검토한 결과, 언어에 따라 포괄하는 구체적인 행동의 범위에 있어서 상당한 차이가 있음을 발견하였으며, 이러한 어의적 차이가 여러 국가들에서 보고하는 괴롭힘 유행률 차이의 상당 부분을 크게 좌우할 수 있음을 지적한 바 있다. 따라서 bullying 개념의 정확한 평가를 위해서는 이 용어가 가지는 문화적 보편성뿐 아니라 이 용어의 번역어가 갖는 문화적 특수성 측면을 고려하여, 이 개념에 포함되는 구체적인 행동(하위 유형)의 범위를 명확히 하는 것이 매우 중요하다 할 수 있다.

적절한 용어 선택

‘bullying’의 개념이 국내에 소개되면서 국내 연구자들 사이에서 이 용어는 ‘괴롭힘’ 외에도 다를 다양한 용어로 번역되어왔다. ‘집단따돌림’(이춘재, 박금주, 1999), ‘왕따’(권준모, 1999), ‘집단괴롭힘’(강진령, 유형근, 2000), ‘또래괴롭힘’(이승연, 2013) 등이 그것이다. 심지어는 ‘학교폭력’이라는 용어와도 동일시되어 학교폭력을 정의하는 요소로 Olweus(1993)의 3대 핵심

요소가 그대로 인용되는 경우도 볼 수 있다(정종진, 2013; 조대훈 등, 2013). 물론 학교에서 발생하는 학생들 간에 발생하는 폭력사건의 경우 괴롭힘인 경우가 많아서 이 둘을 정확히 구분하기 어려운 측면이 없지 않지만(이규미 등, 2014; Batsche, 1997), 개념적으로 괴롭힘과 학교폭력은 구분될 필요가 있다. 첫째는 흉기소지, 방화, 폭행, 자살시도 등 다양한 형태의 폭력 사례가 괴롭힘과 직접적인 관련성 없이 나타날 수 있기 때문이며, 둘째는 ‘학교폭력’에 대한 현재의 법률적 정의(학교폭력 예방 및 대책에 관한 법률의 정의)와 괴롭힘에 대한 학술적 정의를 비교해 보면, 결코 대등한 의미로 해석할 수 없기 때문이다. 학교폭력의 현 법률적 정의는 학교폭력의 대상, 장소 및 결과와 함께 학교폭력에 해당하는 15가지 행위를 명시하고 있는데, 여기에는 형법의 적용대상이 되는 행위뿐 아니라 강제적 심부름, 따돌림, 사이버따돌림 같이 괴롭힘의 하위 유형(Olweus, 2007)에는 포함될 수 있지만 형법의 직접적 적용대상이 되기 어려운 행위도 포함되어 있어서 학교폭력은 적어도 드러나는 행위 자체만으로는 괴롭힘과 상당부분 중복되는 개념으로 이해할 수 있다. 그러나 이러한 정의는 괴롭힘을 다른 형태의 공격행동과 구분해주는 본질적 요소(즉, 의도성, 반복성, 힘의 불균형)를 전혀 반영하고 있지 않다는 점에서 괴롭힘에 대한 정의와 큰 차이를 보인다. 따라서 괴롭힘과 학교폭력이라는 용어가 혼용되는 경향이 있음에도 불구하고 학교폭력은 괴롭힘을 포함하는 보다 포괄적인, 법률적·관습적 용어로 이해되어야 하며(김혜원, 2013; 이규미 등, 2014), 괴롭힘은 학술적으로 보다 잘 정의된 특수한 형태의(앞서 언급한 구체적인 핵심 요소를 갖춘) 학교폭력 혹은 학생들

간의 공격행동으로 개념화하는 것이 합당하다.

한편, 동일한 법률에서 ‘따돌림’은 ‘학교 내외에서 2명 이상의 학생들이 특정인이나 특정 집단의 학생들을 대상으로 지속적이거나 반복적으로 신체적 또는 심리적 공격을 가하여 상대방이 고통을 느끼도록 하는 일체의 행위’로 정의되고 있다. 이러한 정의에서 흥미로운 점은 따돌림의 주요 내용을 사전적인 의미인 “밉거나 싫은 사람을 따로 떼어 멀리하다.”(네이버국어사전, 2014)와 같이 상대방을 배제하고 소외시키는 행위(즉 ‘관계적 공격’에 해당하는)로 보다는 신체적, 심리적으로 공격하고 괴롭히는 행위로 정의하고 있다는 점이다(김혜원, 2013). 이 정의는 또한 괴롭힘의 ‘의도성/유해성’과 ‘지속성/반복성’의 의미를 반영한 것 외에 ‘2인 이상의 학생들’이라는 외형적 요건을 포함시키고 있다. 즉 법률적으로 정의된 따돌림은 괴롭힘을 정의하는 기존의 2대 요소를 언급하고 있다는 점에서 괴롭힘의 학술적 정의에 근접한 것으로 보이지만, 힘의 불균형이라는 핵심 요소를 언급하는 대신 ‘2인 이상의 학생들’이라는 제한 조건을 부과함으로써 괴롭힘과 동일한 용어로 보기에 그 의미가 모호하다. 이러한 따돌림의 정의는 또한 최근의 법학 연구문헌(김종구, 박지현, 2014; 정한중, 2009)에서도 구체적인 유형과 적용범위를 명확히 제시하지 못하고 있다는 비판을 받고 있다.

이렇게 보면, 국내에서 사용되는 괴롭힘(bullying의 번역어로서)이라는 용어는 분명 법률적인 정의나 사전적인 의미의 ‘따돌림’보다는 상위의 개념으로 보아야 하며, 학교폭력의 개념보다는 훨씬 구체적인 특성을 갖춘 공격행동의 하위 유형으로 보는 것이 타당하다. 또한 괴롭힘을 정의하는 Olweus(1993)의 3대

핵심 요소를 기준으로 본다면, 괴롭힘의 관계가 항상 ‘나이나 수준의 유사성’이나 ‘2명 이상의 가해자 집단’을 전제조건으로 성립되는 것은 아니므로(예: 한 명의 상급생이 힘의 불균형 관계를 통해 한 명의 하급생을 지속적, 의도적으로 괴롭히는 경우), 괴롭힘 개념을 본의 아니게 축소하는 것을 막기 위해 ‘또래’나 ‘집단’같은 명사를 붙여서 복합명사화하기 보다는 단순히 ‘괴롭힘’으로 지칭하는 것이 보다 적절할 것으로 보인다. 다만 괴롭힘은 학교 이외의 상황(예: 직장, 사이버 공간)에서도 발생하므로 필요한 경우 ‘학교’ ‘사이버’ ‘직장’ 등의 수식어를 동반하여 표현할 수 있을 것이다.

이상으로 살펴 본 바와 같이, 괴롭힘 본질적 요소를 어떻게 개념화하느냐에 따라, 그리고 괴롭힘에 포함되는 구체적인 공격행동의 범위를 어떻게 규정하고, 어떤 용어를 사용하느냐에 따라 괴롭힘의 정확한 의미는 상당히 달라질 수 있다. 국내 연구자들 간에 가급적 일관성 있고 통일된 용어를 사용하는 것이 효율적인 의사소통을 위해서 필요하겠지만, 보다 중요한 점은 동일한 현상(예: 괴롭힘)을 두고 서로 다른 용어(예: 따돌림, 학교폭력)를 사용하거나, 잠재적으로 서로 다른 현상(예: 괴롭힘과 학교폭력)에 대해 동일한 용어(예: 괴롭힘)를 사용할 경우 이러한 용어들을 해석하여 관련 경험을 보고하는 학생들의 반응결과와 이에 대한 해석이 크게 달라질 수도 있다는 점이다. 이는 괴롭힘 측정과 평가의 타당성과 직결되는 문제인 것이다.

괴롭힘 개념화의 현 쟁점들은 추상적 수준에서 풀어가기 보다는 실제적인 괴롭힘의 측정 문제를 고려하여 조작화 수준에서 고찰해볼 필요가 있다. 따라서 다음 절에서는 괴롭

힘 개념화의 쟁점이 되는 핵심 요소와 하위 유형이 자기보고식 괴롭힘 측정도구에서 어떻게 조작화(operationalization)되고 있으며 관련 쟁점은 무엇인지 구체적으로 살펴볼 것이다.

자기보고식 평가의 쟁점

괴롭힘 연구자들은 면접, 관찰, 교사보고, 또래보고, 자기(학생)보고 등 여러 방법을 활용하여 괴롭힘을 평가해 왔다. 이 중 괴롭힘 연구에서 가장 널리 사용돼 온 것은 자기보고식 평가로, 다른 방법에 비해 실시가 용이하고 편리하며 경제적인 장점으로 인해 가장 선호되고 있다(Sveinsson & Morris, 2007).

괴롭힘의 자기보고식 평가에서는 다음 중 하나의 방식을 통해 괴롭힘을 측정한다. 하나는 학생의 괴롭힘 경험에 대해 묻는 질문을 제시하고 경험여부나 경험빈도를 표기하게 하는 설문조사(surveys) 방식이고, 다른 방식은 괴롭힘을 대표하는 구체적인 행동이나 경험을 표집하여 여러 문항을 구성하고 각 문항의 내용을 경험한 빈도나 문항 내용과 일치하는 정도를 리커트식 척도 상에 평정하도록 하여 각 문항의 값을 합산하는 식의 심리척도(scales) 접근을 사용하는 것이다(Bovaird, 2010; Furlong et al., 2010). 전자의 대표적인 예는 Olweus Bullying Questionnaire(OBQ: Olweus, 2007)로 OBQ의 초판과 개정판은 다양한 언어로 번안되어 여러 나라의 괴롭힘 실태조사에서 많이 사용되고 있다. 국내에서도 몇몇 연구자들(송지윤, 2010; 이춘재, 박금주, 1999; 이해경, 김혜원, 2001)이 OBQ의 주요 문항들을 번안하여 사용한 바 있으며, 현재 교육부(2014)에서 주관하여 한국교육개발원에 위탁하여 실시하고

있는 학교폭력실태조사나 청소년폭력예방재단(2014)에서 실시하고 있는 실태조사는 모두 전자(설문조사)의 방식을 따르고 있다. 후자의 예로는 외국의 Social Experience Questionnaire(Crick & Grotpeter, 1996)나 Reynolds Bully Victimization Scales(Reynolds, 2003) 같은 대표적인 괴롭힘 척도들과 이러한 척도들 혹은 유사한 외국의 척도들을 번안하거나 참조하여 국내 연구자들이 개발한 척도들(김현주, 2003; 서미정, 2008; 최보가, 임지영, 1999; 최은숙, 1999)을 들 수 있다. 다음은 괴롭힘 평가에 대한 최근 연구문헌들을 토대로 자기보고식 평가에 관한 대표적인 쟁점들을 크게 1) 핵심 요소의 조작화, 2) 하위 유형의 조작화, 3) 설문조사에서 단일문항 대 다문항 접근, 4) 실시 맥락과 절차, 5) 응답자의 반응 특성, 6) 공유 방법분산의 문제로 범주화하여 대표적인 연구 결과들과 함께 각 쟁점을 자세히 고찰해 보고자 한다.

핵심 요소의 조작화

다른 형태의 공격행동과 괴롭힘을 구분해주는 핵심 요소를 정확히 측정하는 것은 괴롭힘 평가의 타당도를 확보하기 위해 가장 기초적인 고려사항이다. 이것은 학생들이 경험하는 다양한 공격행동 중 괴롭힘이라는 특정한 형태의 공격행동을 어떻게 조작화하여 측정할 것인가의 문제이다. 이러한 문제에 대한 주요 쟁점을 괴롭힘 용어 및 정의 제시여부 문제, 괴롭힘의 3대 핵심 요소의 조작화 문제로 나누어 고찰하면 다음과 같다.

용어 및 정의 제시 문제

괴롭힘의 핵심 요소를 측정하기 위해 연구

자들이 취하고 있는 가장 보편적인 접근 방식은 자기보고식 설문지에 ‘괴롭힘’이라는 용어와 함께 괴롭힘의 구체적인 정의를 제시하고 그러한 정의에 부합되는 공격행동을 가해자로서 혹은 피해자로서 경험한 적이 있는지를 묻는 것이다. Solberg과 Olweus(2003)는 ‘괴롭힘’이 일상용어 중 하나로 학생들 입장에서 이 용어를 학술적인 의미와 다르게 사용하거나 이 용어의 정확한 의미를 이해하지 못하는 경우가 있을 수 있으므로, 이러한 조치가 괴롭힘 측정의 정확성을 높이는 데 필수적이라고 주장한다. 이를 뒷받침하는 연구로 Vaillancourt 등(2010)의 연구에서 학생들에게 괴롭힘이 무엇인지 정의하도록 했을 때, 거의 모든 학생들(92%)이 괴롭힘이 ‘부정적’ 행동이라는 점을 언급한 반면, 오직 26%의 학생들만이 힘의 불균형 요소를 언급하였다. 같은 맥락에서 Land(2003)의 연구에서도 147명의 중고생에게 괴롭힘 사례를 제시하도록 하였을 때, 절반 이상의 사례에서 반복성의 요소가 나타나지 않았다. 이러한 결과들은 학생들이 이해하는 괴롭힘이 학술적으로 정의된 괴롭힘의 의미와 현저하게 다를 수 있음을 의미하며, 따라서 그러한 차이에서 오는 오차를 줄이기 위해서는 괴롭힘이라는 용어의 정확한 정의를 제시하고 괴롭힘을 측정하는 것이 중요함을 시사한다.

그러나 자기보고식 측정도구에 괴롭힘 용어를 노출하고 그 정의를 제시하는 방식에 대하여 부정적인 입장을 취하는 학자들도 있다. 예컨대, Hamby와 Finkelhor(2000)는 ‘괴롭힘’이라는 용어의 사용 자체가 학생들에게 정서적 부담감을 준다고 주장하였다. Furlong 등(2010)도 ‘괴롭힘을 당했음’을 인정하는 것이 자신의 약점을 인정하는 것이라고 느낄 수 있을 뿐더

러 모든 괴롭힘 피해학생들이 괴롭힘 피해자라는 자기상을 내면화한 것은 아니기에 이러한 표찰에 동의하지 않을 수 있다는 가설을 제기하였다. 이러한 가설을 간접적으로 지지하는 연구로, 성희롱(sexual harassment)의 피해자에 대한 성인 대상의 한 연구(Schneider, Swan, & Fitzgerald, 1997)에서 성희롱의 구체적인 법적 준거에 부합되는 모든 구체적인 경험(‘성희롱’이라는 용어가 포함되지 않은)에 대해서 동의한 많은 피해자들이 질문지의 끝부분에 “성희롱을 경험하였느냐?”라는 최종 질문에 대해서는 동의하지 않은 것으로 나타났다. 또한 Kert, Coddling, Tryon과 Shiyko(2009)의 연구에서 괴롭힘이라는 용어와 함께 Olweus의 정의를 제시받은 학생들의 경우 그렇지 않은 학생들에 비해 괴롭힘 가해 경험을 유의하게 적게 보고하는 것으로 나타났으며, Vaillancourt 등(2010)의 연구에서도 보다 구체적인 정의가 제시될 경우 괴롭힘 피해 경험을 적게 보고하는 경향이 있었다. 이러한 연구결과들은 대체로 괴롭힘 정의 제시 여부가 괴롭힘 유병률 같은 측정 결과에 의미 있게 영향을 미치는 변인임을 보여준다. 정의 제시에 반대하는 입장에 있는 학자들은 이에 대한 대안으로 심리적도 방식을 선호한다. 즉, 괴롭힘이라는 용어를 아예 사용하지 말고 대신 괴롭힘과 관련된 구체적인 행동들을 제시하고 이러한 구체적 ‘행동’에 대한 동의 정도의 평균치나 합산점수를 구해 극단점수에 있는 응답자를 괴롭힘 가해나 피해자로 간주함으로써 표찰과 관련된 편향을 피할 수 있다고 본다.

괴롭힘에 대한 정의를 자기보고식 측정도구에 제시하는 방식을 택하든 그렇지 않든, 결국 쟁점의 핵심은 괴롭힘을 정의하는 핵심 요소를 어떻게 조작화하여 정확히 측정하느냐의

문제에 있다. 따라서 다음은 각 핵심 요소 별로 조작화 관련 쟁점을 보다 구체적으로 검토하고자한다.

의도성의 조작화

괴롭힘의 핵심 요소 중 하나는 의도성인데 연구자들은 이 요소에 대한 설명을 괴롭힘 정의에 포함시키거나 개별 문항내용에서 가해학생의 의도성이 드러나도록 문항을 작성함으로써 의도성 측정의 문제를 다루어왔다. 가령, OBQ의 경우 괴롭힘을 정의하면서 “우호적이거나 장난스러운 방식으로 놀리는 것은 괴롭힘이 아니다.”라는 부연설명을 제시하는가하면, 다른 도구들의 경우 “고의로”, “심술로” 등의 문구를 넣기도 한다. 그러나 의도성의 조작화 문제는 그렇게 단순하지 않다. 가령, Greif와 Furlong(2006)은 의도성을 가해학생의 실제 의도로 보아야 하는 것인지, 아니면 가해학생의 의도에 대한 피해학생의 귀인으로 보아야 하는 것인지 의문을 제기하였다. 이들은 항상 일치하지는 않으며, 자기보고식 도구를 통해 피해경험을 측정할 경우 불가피하게 가해학생의 실제 의도보다는 피해학생의 귀인에 따라 의도성이 측정된다고 할 수 있다. 문제는 피해학생의 오귀인(misattribution)이 피해경험 보고에 영향을 줄 수 있다는 점인데, 괴롭힘 가해/피해 중복경험 집단이 가해집단, 피해집단 및 무경험 집단에 비해 모호한 상황에서 의도성 귀인성향이 높다는 연구결과(Camodeca, Goossens, Schuengel, & Meerum, 2003)는 괴롭힘 피해경험이 이러한 요인에 의해 편향(과잉)보고될 가능성을 시사한다.

반복성의 조작화

괴롭힘의 자기보고식 도구들은 대부분 반복

성/지속성의 요소를 명시하는 문구를 정의 부분에 넣거나 특정한 기간 동안 구체적인 괴롭힘 관련 행동을 얼마나 자주 실행 혹은 경험하였는지를 표기하도록 함으로써 반복성 측정의 문제를 다룬다. 이러한 측정 방식과 관련하여 현재 두 가지 주요 쟁점이 존재한다. 첫 번째 쟁점은 괴롭힘 관련 경험의 자기보고 시 적절한 ‘시간참조틀(time frames of reference)’(예: ‘지난 주’, ‘지난 달’, ‘지난 학기’, ‘지난 해’ 등)은 무엇이고, 괴롭힘 경험빈도에 따라 범주화(예: 가해자 집단, 피해자 집단, 가해-피해자 집단, 무경험 집단 등)하여 분류할 때 사용할 수 있는 적절한 절단점(cut-off points)은 무엇인가(예: ‘1주에 여러 번’, ‘1주에 한 번’, ‘1달에 2~3번’ 등)에 관한 것이다. 현재 여러 자기보고식 도구에서 사용하는 적절한 시간 참조틀과 절단점에 대해서는 다양한 견해가 있다. 지난 1달 동안의 경험(즉, 최근의 경험)을 묻는 것이 지난 1년 동안의 경험(즉, 먼 과거의 경험)을 묻는 것보다 정확한 결과를 산출한다는 주장(Furlong et al., 2010)이 있는가 하면, 지난 1달 동안의 경험을 묻는 경우 지난 1년간의 경험을 묻는 경우에 비해 보다 덜 심각하지만 빈번히 발생한 사건의 회상을 촉진하여 괴롭힘 피해 관련 경험을 오히려 더 많이 보고하는 경향이 있음을 보여주는 연구결과(Morrison & Furlong, 2002)도 제시되었다. 또한 23개의 대표적인 자기보고식 도구를 비교 분석한 Swearer 등(2010)의 연구에 따르면, ‘가끔’이나 ‘자주’ 같이 모호하게 정의된 절단점을 사용하는 경우에서부터 ‘한두 번’에서부터 ‘1주에 1회 이상’에 이르기까지 다양한 값이 절단점으로 사용되고 있어서 반복성을 어떻게 조작화할지 연구자들 간에 합의가 이루어지지 않고 있는 것을 잘 알 수 있다. 다만, 적절한

절단점과 관련해서 Solberg와 Olweus(2003)는 ‘1달에 2~3번 이상’의 괴롭힘을 경험하는 학생들의 경우 그렇지 않은 학생들에 비해 우울, 자존감, 또래에 의한 수용감 등에 있어서 유의한 차이를 보인다는 연구결과에 기초하여 이러한 빈도를 절단점으로 사용할 것을 권장한 바 있다.

반복성의 측정과 관련한 두 번째 쟁점은 괴롭힘으로 간주하기 위해서는 동일한 괴롭힘이 동일인에 의해 반복되어야 하는지, 아니면 동일한 학생을 다수의 학생들이 일회적으로 괴롭히는 경우도 반복성 요건을 충족시키는 것으로 보아야 하는지에 관한 것이다. Greif와 Furlong(2006)은 현재의 측정도구들이 반복성의 이러한 측면을 전혀 다루고 있지 않음을 지적하면서, 한 가해학생의 강도 높은 반복적 괴롭힘이 다수 학생들의 비교적 사소한 괴롭힘에 비해 더 유해한 것인지를 답하기 위해서는 동일 학생에 의한 반복성뿐 아니라 다수의 학생에 의한 반복성 또한 구분하여 측정되어야 한다고 주장하였다. 요컨대, 현재 반복성의 조작화는 측정도구에 따라 다양한 방식으로 이루어지고 있으며, 일부 학자들이 권장사항을 제시하는 정도에 그치고 있어서 괴롭힘 측정치의 신뢰도, 타당도 및 연구 간의 비교가능성을 저해하는 주요 요인이 되고 있다고 할 수 있다.

힘의 불균형 조작화

피해학생과 가해학생 간에 존재하는 힘의 불균형은 괴롭힘 역동이 쉽게 변화하지 않는 중요한 이유이며, 괴롭힘을 다른 형태의 공격행동과 구분 짓는 핵심 구성요소다(Espelage & Swearer, 2003; Greif & Furlong, 2006). 의도성이나 반복성의 경우와 흡사하게 괴롭힘의 정의

를 제시하는 OBQ 같은 자기보고식 도구는 “비슷한 힘을 가진 두 학생이 싸우거나 다투는 경우는 괴롭힘이 아니며...괴롭힘을 당하는 학생은 자신을 방어하기 어렵다.”(Solberg & Olweus, 2003, p. 246)는 조건을 제시함으로써 힘의 불균형 요소가 반영된 공격행동을 측정하고자 시도한다. 그러나 단지 기계적으로 적절한 정의를 제시하는 것만으로 이 요소의 측정을 보장할 수 없음을 시사하는 연구들이 보고되고 있다. 예컨대, Baly와 Cornell(2011)은 여러 중학생 학급을 실험조건과 통제조건에 무선할당한 후, 실험조건에 할당된 학급의 경우 괴롭힘에 내재된 힘의 불균형 요소를 강조하는 6분짜리 교육용 비디오를 본 후 괴롭힘에 대한 Olweus식의 정의를 읽고 괴롭힘 경험 조사를 기입하게 하였고, 통제조건의 경우 동일한 정의가 포함된 조사를 기입하되 다만 비디오를 보지 않은 채 기입하도록 하였을 때, 실험조건의 학생들이 통제조건의 학생들에 비해 사회적 괴롭힘의 경우 무려 48%, 신체적 괴롭힘의 경우 무려 54%나 더 적게 괴롭힘 경험을 보고한다는 결과를 보고하였다. 더욱이, 괴롭힘이라는 용어를 사용하지 않고 괴롭힘의 정확한 정의를 제시하지 않는 심리척도 방식의 자기보고식 도구들의 경우, 힘의 불균형은 거의 무시돼온 요소다(Swearer et al., 2010). Swearer 등(2010)은 구체적인 용어 정의를 제시하지 않고 괴롭힘을 측정하는 7개의 대표적인 괴롭힘 측정 도구들을 비교분석하면서, 7개의 도구 모두가 힘의 불균형 요소를 적절히 측정하고 있지 않으며, 이 중 5가지는 지속성이나 의도성 중 하나만을, 2가지는 이 두 요소를 측정하고 있다고 보고하였다. 같은 맥락에서 Furlong 등(2010)도 괴롭힘 피해자-가해자 관계에서 힘의 불균형을 보고한 피해학

생들이 그렇지 않은 피해학생들이나 괴롭힘 피해경험이 없는 학생들에 비해 유의하게 취약한 정신건강 상태를 보인다는 증거(You et al., 2008)를 제시하면서 자기보고식 괴롭힘 척도 제작 시 힘의 불균형이 요소가 포함되도록 괴롭힘 측정의 정밀성을 높일 것을 촉구하였다. 이러한 문제점은 국내 연구에서 사용되고 있는 대표적인 척도들(김현주, 2003; 최은숙, 1999; 서미정, 2008; 최보가, 임지영, 1999)을 검토해 보아도 동일하게 발견되고 있다.

요컨대, 자기보고식 평가에서 괴롭힘 핵심 요소의 조작화 문제는 현재로서 일관성이나 합의보다는 다양성을 보이고 있다. 반복성의 조작화와 관련하여 시간참조틀 및 절단점 설정에 대해 일관된 기준을 사용하는 것은 연구들 간의 비교가능성 및 진단과 분류의 타당도를 높이기 위해 꼭 해결해야 할 과제인 것으로 보이며, 특히 힘의 불균형은 괴롭힘의 현 개념화와 조작화 간에 가장 큰 괴리를 보여주는 요소로, 괴롭힘 측정의 정밀성과 타당성을 확보하기 위해서 반드시 개선되어야 할 문제로 보인다.

하위 유형의 조작화

많은 괴롭힘 연구자들은 신체적, 언어적 괴롭힘 같은 직접적 형태의 괴롭힘 외에도 간접적 형태의 괴롭힘(사회적 혹은 관계적 괴롭힘), 성적 괴롭힘, 최근에는 사이버괴롭힘(cyberbullying)까지 그 범위를 확대하여 괴롭힘을 개념화하고 있다. 괴롭힘의 하위 유형이 확장됨에 따라 보다 다양한 공격행동이 괴롭힘으로 인식되고, 따라서 괴롭힘 유병률 측정에 중요한 변수로 작용할 수 있다. 그러나 하위 유형을 지나치게 확대하여 괴롭힘을 넓게

정의하는 것이 개념적 혼선을 초래할 수 있으며, 자료 분석 상의 난점을 야기하므로 비생산적이라는 주장이 제기되기도 하였다(Cornell & Cole, 2012; Sveinsson & Morris, 2007). 예를 들면, Cornell과 Cole(2012)은 다양한 괴롭힘 행동을 단일 개념 속에서 측정함으로써 하위 유형이 지닌 독특한 특성이 묻혀 버릴 수 있다는 우려를 제기하였으며, 괴롭힘 유병률 연구에서 많은 하위 유형별로 빈도, 강도, 지속성 등을 별도로 측정하고 가해학생, 피해학생, 가해-피해학생, 무경험 집단 등으로 분류하는 것은 자료 분석 상의 과도한 복잡성을 초래할 뿐 아니라 대부분의 연구에서 표본 크기의 문제를 제기한다고 지적하였다. 이러한 쟁점은 하위 유형의 추가가 괴롭힘 측정 결과에 미칠 수 있는 영향을 충분히 고려하여 연구에 포함될 적절한 하위 유형을 신중히 결정해야 함을 의미한다.

설문조사에서 단일문항 대 다문항 접근

자기보고식 괴롭힘 설문조사에서 단일문항으로 괴롭힘 가해 및 피해경험을 측정하는 것과 여러 문항으로 측정하는 것은 어떠한 차이가 있는가? 단일문항 측정은 전수조사 같은 대규모 평가에서 비교적 쉽게 괴롭힘 실태를 파악할 수 있는 장점이 있지만, 심리측정적 관점에서 본다면 다문항으로 구성된 측정도구에 비해 신뢰도가 낮을 가능성이 크므로 적합한 측정 문항 수에 대한 결정은 쟁점이 될 수 있다. 예컨대, Solberg와 Olweus(2003)는 유병률 추정을 연구목적으로 하는 경우 잘 정의된 단일문항을 구체적인 반응선택안과 함께 제시함으로써 결과의 해석, 활용 및 연구 간의 비교를 용이하게 할 수 있다고 제안한 반면,

Sharkey, Dowdy, Twyford와 Furlong(2012)은 괴롭힘을 포함한 학교폭력 관련 실태조사에서 제시하는 추정치들의 신뢰도 문제를 제기하면서 다문항을 사용하는 심리측정적 접근을 촉구하였다.

OBQ 같은 대표적인 자기보고식 설문지의 예를 보면, 단일문항 접근과 다문항 접근을 동시에 따르고 있다. 가령, 괴롭힘의 정확한 정의를 제시하고, “지난 2~3개월 동안 괴롭힘을 당한 적이 있는가?”를 전반적으로 물음으로써 단일문항으로 괴롭힘 피해경험을 측정하고, 이후 괴롭힘의 하위 유형은 구체적인 내용을 담고 있는 다문항 형태로 제시된다. 최근 연구(Vaillancourt et al., 2010)에 따르면, OBQ에서 사용되는 전반적 단일문항 질문이 괴롭힘 경험이 실제로 전혀 없는 학생들을 비교적 쉽게 가려내는 데는 효과적이지만, 이러한 질문에 ‘아니오’로 응답한 학생들 중 일부는 이후에 제시되는 구체적인 질문에서 ‘예’로 반응하는 경우가 있어서 구체적인 다문항 접근이 실제 괴롭힘을 경험하는 학생들을 탐지하는 데 보다 민감한 것으로 나타났다. 그러나 현재로서 다문항 접근을 사용하는 경우 좀더 정확한 결과를 산출하는 것인지, 아니면 과잉보고를 유도하는 것인지 여부는 확실하지 않으며, 따라서 외적 준거를 통해 타당도를 검토하는 후속연구를 통해 확인될 필요가 있다.

평가 맥락과 절차

자기보고식 평가를 실시하는 절차 혹은 맥락과 관련된 여러 요인들이 괴롭힘 측정 결과에 영향을 미치는 것으로 알려져 있으며, 이러한 요인들을 어떻게 통제하여 괴롭힘을 측

정할 것인지의 문제는 주요 쟁점 중 하나다. 이러한 요인들 중 익명성, 설문조사 실시자의 특성, 설문조사의 형태와 관련된 쟁점을 살펴보면 다음과 같다.

익명성 대 비밀보장 실시

괴롭힘 관련 경험의 솔직한 보고를 촉진하기 위해 괴롭힘 설문조사는 흔히 익명(anonymity)으로 실시되며, 또한 이러한 절차가 권장되기도 한다(Solberg & Olweus, 2003). 그러나 익명으로 실시하는 설문조사는 자기보고 결과의 타당성을 외적 준거를 통해 평가할 기회를 제한하는 문제점이 있다. 즉 괴롭힘 설문조사에 준거로 사용할 변인을 동시에 포함시키지 않는 한, 이후에 얻은 자료 등과 관련지어 자기보고의 정확성을 독립적으로 평가할 방법이 없는 것이다. 그렇다면 익명성을 보장할 때 학생들은 과연 괴롭힘을 보다 정확하게 보고하는가? 이 질문에 대해서는 여전히 논쟁이 계속되고 있다. 관련 실증연구들은 대체로 익명성과 괴롭힘 보고율 간의 관계를 연구하였는데 서로 상치되는 결과를 보고하고 있다. 예컨대 Ahmad와 Smith(1990)는 익명으로 실시하는 자기보고식 도구(OBQ)에서 괴롭힘 피해경험을 보고한 학생들 중 85%만이 (익명보장이 불가능한) 면접에서도 피해경험을 인정하였으며, OBQ에서 가해경험을 보고한 학생들 중 오직 50%만이 면접에서 가해경험을 인정하였다는 결과를 제시하였다. 반면, Chan, Myron과 Crawshaw(2005)는 괴롭힘 설문조사지를 익명으로 실시하는 조건이나 기명으로 실시하는 조건에 학생들을 무선할당하여 괴롭힘 가해 및 피해경험의 보고율을 비교하였는데, 두 조건 간에 유의한 차이를 발견하지 못했다. Chan 등(2005)의 연구는 ‘괴롭힘’이라는 용어를

제시하지 않았다는 점에서 Ahmad와 Smith (1990)의 연구와 차이가 있어서, 두 연구를 직접 비교하기는 어렵다고 할 수 있으며, 익명성이 제3의 변수와 상호작용하여 보고율에 영향을 미치는지 추가 연구가 필요하다.

한편, 다른 연구자들은 익명성 대신 ‘비밀보장(confidentiality)’의 조건에서 괴롭힘 경험을 측정하는 방법이 오히려 괴롭힘 경험을 정확히 측정하는 대안이 될 수 있다고 주장한다. 이를 지지하는 실증연구로, O'Malley, Johnston, Bachman과 Schulenberg(2000)는 익명성 대 비밀보장의 조건 하에 무선할당된 학생들이 약물 사용이나 불법적 행동(도박, 무기소지 등)과 같은 민감한 정보를 보고하는 비율에 있어서 서로 유의한 차이를 보이지 않는다고 보고하였다. 이 연구는 괴롭힘 보고율을 직접 다른 연구가 아니라는 점에서 일반화의 한계점이 있지만, 학생들이 학급풍토가 지지적이며 교사들이 효과적으로 개입할 것이라는 신념을 가질수록 괴롭힘 상황에서 교사의 도움을 요청할 가능성이 크다는 연구결과(Eliot, Cornell, Gregory, & Fan, 2010)에 비추어 볼 때, 비밀보장의 확신이 있는 상황이라면 괴롭힘 경험을 과소보고할 가능성은 줄어들 것으로 유추해 볼 수 있다.

그런데 익명성 보장이 괴롭힘 보고율을 증가시키는지의 여부와 상관없이, 괴롭힘 보고율이 높다고 해서 이러한 결과가 반드시 정확한 보고임을 의미하지는 않는다는 점에 주목할 필요가 있다. Cornell과 Bandyopadhyay(2010)는 앞서 제시한 Ahmad와 Smith(1990)의 연구결과를 재해석하면서, 자기보고식 도구에서는 부주의하게 응답하였던 일부 학생들이 면접에서는 보다 신중하게 응답하였기 때문에 보고율에 차이가 발생했을 가능성을 제기하였다.

요컨대, 자기보고식 도구에서 익명성 보장은 기명 방식에 비해 괴롭힘 경험의 보고율을 증가시킬 가능성이 있지만 높은 보고율이 반드시 정확한 보고를 의미하지는 않으므로 익명성 대 비밀보장 조건에서 나타나는 자기보고 결과의 정확성을 외적 증거에 비추어 비교하는 실증연구가 절실히 필요하다고 할 수 있다.

설문조사자의 특성 및 설문조사 형태

괴롭힘의 자기보고식 평가에서 익명성 같은 맥락 요인 외에도 조사실시자 대상의 훈련 제공 여부나 설문조사를 전통적인 지필검사 형식으로 실시할지 아니면 다른 대안적인 형태로 실시할지의 여부 또한 쟁점이 되고 있다. 예컨대, Cross와 Newman-Gonchar(2004)는 학생들을 대상으로 설문조사를 실시하는 책임을 맡은 일군의 교사들을 대상으로 간단한 훈련을 제공하였는데, 훈련 내용은 학생들에게 조사자료의 용도나 중요성을 설명하고, 정직한 답변을 촉구하는 설명을 제공하도록 돕는 것이었다. 이러한 훈련을 받은 교사들과 그렇지 않은 교사들의 책임 하에 실시한 조사자료의 결과를 비교하였을 때, 훈련을 받은 교사들을 통해 수집한 자료의 경우 무효반응의 비율(3%)이 다른 조건에서 수집한 자료의 경우(28%)에 비해 현저하게 낮은 것으로 나타났다. 이러한 결과는 괴롭힘 설문조사 시 비교적 간단한 교사훈련을 통해 학생들의 자기보고 결과의 신뢰도를 높일 수 있음을 보여준다.

한편, Turner 등(1998)은 무기소지, 폭력행위나 위협 등 개인적으로 민감한 경험을 묻는 자기보고식 설문조사(괴롭힘은 포함되지 않음)를 전통적인 지필 형태로 실시할 때와 컴퓨터를 활용하여 실시할 때 학생들의 경험 보고율에서 차이가 있는지 검토하였다. 연구결과 후

자의 경우 유의하게 높은 보고율을 보였다. 또한 Hilton, Harris와 Rice(2003)의 연구에서는 청소년들의 폭력 가해 및 피해경험을 조사하기 위해 문항을 지필 형태로 제시하거나 녹음하여 제시한 후, 경험 보고율에 차이가 있는지 비교하였는데, 놀랍게도 지필검사 형태로 제시된 경우 오디오로 제시된 경우에 비해 2~3배나 더 많은 경험을 보고하였다. 이러한 연구결과들은 전통적인 지필조사 형태가 대안적인 형태에 비해 더 정확한지 아닌지에 대한 답변은 제시하지 못하지만, 명백히 조사 형태와 외적 준거를 관련짓는 타당화 연구가 필요함을 시사하고 있다고 할 수 있다.

응답자의 반응 특성

학생의 자기보고에 기초한 괴롭힘 평가는 응답자가 '보고한 경험'을 측정한다는 점에서 응답자 특유의 반응 특성에 의해 영향을 받을 수 있으며, 이러한 특성은 측정의 오차분산을 초래할 수 있다. 따라서 보고된 경험과 학생의 실제 경험 간의 차이를 초래하는 응답자의 반응 특성에 주목하고 이러한 간극을 어떻게 좁힐 수 있는지의 문제는 괴롭힘 측정에 있어서 중요한 도전이며 쟁점이다. 자기보고 결과에 영향을 미치는 것으로 밝혀진 응답자 반응 특성 관련 요인들로는 무효반응, 민감도 증가 및 요구특성의 문제가 있다.

무효반응

괴롭힘을 자기보고식 도구로 측정할 경우 일부 학생들은 진지하게 응답하기 보다는 부주의하게 응답하거나 의도적으로 왜곡하여 보고할 수 있다. 이러한 응답자들은 문항의 내용과 상관없는 동일반응, 무선반응, 극단반응

혹은 패턴화된 반응을 보일 수 있으며 이러한 반응들을 최소화하고 적절히 걸러내는 절차를 거치지 않는다면, 괴롭힘의 측정 결과는 크게 왜곡되거나 편향될 수 있음을 시사하는 연구 결과들이 제시되었다. 일례로 만 명 이상의 청소년을 대상으로 고위험 행동에 대한 설문 조사를 실시한 Cornell과 Loper(1998)의 연구에서, 잠재적인 무효반응을 선별하는 기준을 수립하여 '의심쩍은' 응답자를 선별하여 제외시켰을 때, 대부분의 고위험 행동의 유병률 추정치가 작게는 3%에서 많게는 10%까지 감소함을 보고하였다. 유사하게, Cross와 Newman-Gonchar(2004)에서도 연구표본에 포함된 한 고등학교의 경우 괴롭힘 피해율이 46%로 조사되었지만, 일정한 기준에 의해 일정한 무효반응 선별절차를 거쳐 다시 산정한 결과 25%로 감소되는 결과를 보고한 바 있다. 이러한 두 연구결과를 논의하면서 Cornell과 Cole(2012)은 자기보고식 도구에서 이러한 무효반응을 제거함으로써 감소되는 괴롭힘 유병률의 크기가, 많은 괴롭힘 예방 프로그램을 통해 감소되는 유병률 크기보다 더 크다고 지적하였다.

민감성 증가 및 요구특성

괴롭힘 연구에서 예방 프로그램 실시에 따른 괴롭힘 유병률의 변화를 측정하는 경우가 많이 있는데, 이러한 변화를 자기보고식 도구를 통해 평가할 경우 민감성 증가와 요구특성의 문제는 평가의 정확성을 잠재적으로 위협하는 요소로 고려될 필요가 있다. 예컨대, Smith와 Ananiadou(2003)는 괴롭힘 예방프로그램이 괴롭힘 보고에 대한 학생들의 민감성을 증가시켜 자기보고식 평가에서 더 많은 괴롭힘 경험을 보고하도록 유도하는 경향이 있다고 주장하였다. 이는 괴롭힘 프로그램이 사실

상 효과적일수록 학생들은 괴롭힘 경험을 더 많이 보고할 가능성이 있다는 점을 의미한다. 이러한 주장을 뒷받침하는 연구로, Frey 등 (2005)의 연구에서 괴롭힘 예방프로그램 실시 후 괴롭힘 관련 행동을 훈련된 관찰자가 평가하였을 때는 괴롭힘이 감소한 것으로 나타났지만, 학생들의 자기보고 결과에서는 변화가 없는 것으로 나타났다. 이러한 민감성 증가와는 반대 방향으로 작용하여 민감성 문제를 더욱 복잡하게 만드는 것은 괴롭힘 예방프로그램이 진행됨에 따라 학생들이 괴롭힘 경험을 연구자가 기대되는 방향으로 왜곡하여 보고하는 문제다(Cornell & Bandyopadhyay, 2010). 즉 괴롭힘의 부적절함에 대해서 학생들이 반복적으로 교육을 받음에 따라, 일부 학생들은 괴롭힘 경험의 실제 감소와는 상관없이 연구자의 기대에 목중함으로서 괴롭힘 경험의 감소를 보고할 가능성이 있다는 점이다. 따라서 한 연구에서 이러한 요구특성의 문제와 민감성 증가의 문제가 체계적으로 작용한다면, 예방프로그램이 적용된 초기에는 괴롭힘의 자기보고결과가 증가한 것으로 나타났다가, 점차 프로그램의 요구특성에 맞추어 반응함에 따라 괴롭힘이 감소한 것으로 보고하는 패턴이 나타날 수 있다(Cornell & Cole, 2012).

측정도구의 공유방법분산 문제

동일한 응답자를 통해 얻은 두 측정치가 상호 관련된 것으로 나타날 때, 관찰된 공분산이 두 측정치가 대표하는 구인의 실제 관계보다는 반응방식의 일관성 혹은 측정방법의 동일성에 의해 기인한다면 이는 두 변인 간의 관찰된 관계가 체계적인 측정오차에 기인하는 것으로 보아야 하며 이를 공유방법분산(shared

method variance) 혹은 단일방법편향(monometric bias)이라고 부른다(Cornell & Cole, 2012). 이러한 문제는 자기보고식 평가에서 다양한 요인, 예컨대 학생의 독해력, 기분, 수검 태도 등에 의해 초래될 수 있으며, 측정결과를 오염시키거나 변인 간의 실제 관계를 과장하는 결과를 초래한다(Cornell & Bandyopadhyay, 2010). 이러한 오차는 괴롭힘 연구에서 흔히 간과되거나 평가절하 되어왔지만 일반적으로 행동과학 분야에서 심각한 문제로 인정되고 있으며(Podsakoff, MacKenzie, Lee & Podsakoff, 2003), Cote와 Buckley(1987)는 전형적인 행동과학 연구에서 측정방법의 동일성에 기인하는 체계적인 측정오차가 총 분산의 1/4 정도라고 추정하기도 하였다.

공유방법분산이 괴롭힘 연구에서 특별히 쟁점이 되는 이유는, 괴롭힘 평가가 흔히 익명의 자기보고식 도구를 활용하여 이루어지는 바, 자기보고식 괴롭힘 측정치와 다른 자기보고식 측정치 간의 관찰된 상관관계가 변인 간의 실제 관련성 외에도 측정방법의 동일성에 의해, 정확히 알 수 없는 정도로 과장 혹은 왜곡되었을 가능성이 있기 때문이다. 또한 OBQ 같은 대표적인 측정도구조차도 내적합치도 지수를 보고하거나, 동일방법 측정치 간의 상관을 보고하는 동시타당도 연구에 그치는 등 자기보고식 괴롭힘 측정도구의 적절한 타당화 작업이 아직 제대로 진행되고 있지 못하다는 비판을 받고 있기 때문이다(Cornell & Cole, 2012; Solberg & Olweus, 2003). 이러한 상황은 국내의 경우 더욱 심각하다. 단적으로, 1980년대 이후 진행된 국내 괴롭힘 연구의 경향을 분석한 한 연구(정옥분, 정순화, 김경은, 박연정, 2008)에 따르면, 괴롭힘을 측정하기 위해 연구자가 자가제작하거나 다른 척도들을

임의로 재구성하여 제작한 자기보고식 도구를 사용한 경우가 가장 빈번(47.8%)하였고, 이러한 연구들은 대부분 신뢰도를 보고하지 않거나 내적합치도 계수만을 제시하는 정도에 그쳤으며, 타당도의 경우에도 탐색적 요인분석 결과를 제시하거나 동시타당도의 증거를 제시하는 극소수의 연구를 제외하고는 대부분의 연구에서 측정도구의 심리측정적 속성이 무시되고 있는 것으로 나타났다. 이러한 문제는 청소년폭력예방재단이나 교육부의 실태조사의 경우에도 이 조사들에서 사용되는 자기보고식 도구들의 신뢰도 및 타당도가 현재 전혀 알려져 있지 않았다는 점에서 동일하게 해당된다고 할 수 있다.

이상으로 괴롭힘의 자기보고식 측정과 관련된 조작화 및 방법론적 쟁점들을 6가지로 나누어 제시하고 관련 연구문헌을 통해 각 쟁점을 자세히 고찰하였다. 이러한 쟁점들은 괴롭힘 연구에서 현재 자기보고식 평가가 매우 보편적임에도 불구하고 측정 결과의 신뢰도 및 타당도에 영향을 미치는 여러 난제를 안고 있음을 보여준다. 따라서 괴롭힘 연구의 과학적 진보를 위해서는 무엇보다도 이러한 문제들을 해결하기 위한 노력이 선행되어야 함이 자명해진다.

시사점 및 개선방안

본 절에서는 앞서 분석 제시한 개념화 및 자기보고식 평가 관련 쟁점들이 국내의 괴롭힘 연구자들에게 구체적으로 시사하는 점은 무엇이며, 이에 비추어 자기보고식 괴롭힘 평가의 향후 개선방안은 어떠한지 제언하고자 한다. 논의의 편의 상 개념화 및 조작

화 측면과 측정 전략적 측면으로 나누어서 시사점과 개선방안을 논의하면 다음과 같다.

개념화 및 조작화 측면

첫째, 괴롭힘은 학생들 사이에서 나타나는 다양한 형태의 공격행동 중 특수한 요소를 갖춘 공격행동으로, 개념적 수준에서 학교폭력 같은 보다 포괄적인 용어와는 구분되어야 하며, 이 용어와 종종 혼용되며 의미 상 중첩되는 부분이 있지만 포괄성이나 구체성 면에서 차이가 있는 따돌림 같은 다른 개념과도 구분되어야 한다. 괴롭힘 개념을 유사개념과 차별화하지 않음으로써 괴롭힘이라는 용어는 학교폭력 관련 현상들을 지칭하기 위해 사용되는 다른 용어들과 계속해서 혼돈 및 혼용될 것이며, 학술적 개념인 괴롭힘에 대한 연구결과를 교육현장의 실무자들과 정확히 소통하는 데 걸림돌이 될 것이다. 앞서 살펴보았듯이 괴롭힘의 개념화와 관련하여 의도성과 힘의 불균형은 괴롭힘을 다른 형태의 공격행동과 구분 짓는 핵심 요소로서 현재 가장 널리 동의되고 있는 것으로 보이며, 반복성/지속성 요소가 반드시 필요한지, 그 외의 다른 추가 요소가 필요한지의 여부는 학자들에 따라 이견이 존재하는 것으로 보인다. 따라서 괴롭힘 연구자들은 적어도 이 두 가지 요소를 중심으로 괴롭힘과 다른 관련 개념을 구분하고 이러한 요소를 충실히 반영하여 괴롭힘을 측정하도록 측정방법이나 전략을 개선할 필요가 있으며, 반복성 요소는 현재 대부분의 자기보고식 평가 도구에서 직·간접적으로 측정되는 요소(즉, 이 요소를 정의에 포함시키거나, 경험빈도를 구체적으로 물음으로써)인 만큼 각 연구에서 반복성에 대한 구체적이고 명확한 조작적 정

의를 바탕으로 이 요소의 반영 여부와 방법을 결정하는 것이 최선일 것으로 보인다.

둘째, 괴롭힘의 명확한 개념화는 신뢰롭고 타당한 측정을 위한 필수조건이므로 무엇보다도 연구자들은 괴롭힘의 개념을 조작화 수준에서 최대한 합의하고 이에 기초하여 개발된 설문조사지나 척도를 사용하여 측정 결과를 비교, 분석 및 해석해야 할 필요성이 있다. 이는 괴롭힘 유병률을 측정하는 연구들에서 특히 중요하다고 할 수 있는데, 괴롭힘 유병률을 연구한 22개국 82개 연구들에 대한 최근의 통합분석 연구(Cook, Williams, Guerra, Kim & Sadek, 2010)는 괴롭힘의 구체적인 측정방법론적 특성(예: 자기보고, 또래평가 혹은 교사평가, 시간참조틀, 괴롭힘 정의의 제시 여부)에 따라 유병률이 크게 차이가 남을 보여주고 있다. 이러한 연구결과는 학교폭력 유병률을 측정하기 위해 정규적으로 실시하고 있는 국내의 실태조사에 시사하는 바가 크다. 특히, 실태조사에서 보고되는 피해, 가해응답률이 큰 차이를 보이는데(예컨대, 2014년에 보도된 교육부의 실태조사 자료에서는 학교폭력 피해율과 가해율은 각각 1.4%와 0.6%로 보고된 반면, 비슷한 시점에 보고된 청소년폭력예방재단의 자료에서는 각각 6.1%와 5.7%로 보고된 바 있다.), 둘 다 자기보고식 방법을 사용하였지만 두 조사에서 사용하는 용어나 시간참조틀(예: ‘지난 1년’ 대 ‘지난 6개월’), 경험 빈도를 조작화하는 방식(예: ‘1달에 1~2회’같은 구체적인 빈도 대 ‘보통’ 혹은 ‘자주’ 등의 모호한 용어)의 차이로 인해 유병률이 다르게 측정되었을 가능성을 배제할 수 없다. 괴롭힘의 자기보고식 평가에서 연구자 간의 개념 정의 및 조작화의 차이로 인해 발생하는(알 수 없는 만큼의) 분산을 줄일 수 있는 만큼 연구

간의 정확한 비교가 가능해지며 괴롭힘 연구의 과학적 진보를 촉진할 수 있으므로 연구자들은 이러한 요인들에 좀 더 세심한 주의를 기울일 필요가 있다.

셋째, 자기보고식 도구에서 괴롭힘이라는 용어가 측정도구에 직접 사용되는 경우 괴롭힘의 의미를 학생들이 정확히 이해한 상태에서 응답하도록 돕는 조치가 필요하다. 괴롭힘에 대한 적절한 정의 없이 괴롭힘이라는 용어를 사용하여 단순히 학생들의 경험 빈도를 묻는 방식은 연구자가 정의하는 괴롭힘 보다는 학생들이 ‘이해한’ 괴롭힘을 측정할 가능성이 높다는 점에 유의 할 필요가 있다. 이는 현재 학교폭력 실태조사에서 흔히 사용하는 방식(즉, 학교폭력의 하위 유형으로서 ‘괴롭힘’과 ‘따돌림’ 경험을 이들 용어에 대한 명확한 정의를 제시하지 않은 채 단순히 경험여부나 빈도만을 묻는 형식)이 대체로 적절하지 못함을 의미한다. 괴롭힘이라는 용어를 사용한다면 용어와 함께 명확한 정의를 제시할 필요가 있으며(Solberg & Olweus, 2003), 무엇보다도 학생의 발달수준에 맞추어 적절한 문장이나 표현을 사용하여 학생의 이해를 확보하도록 하는 것이 제일 중요하다고 할 수 있다. 이와 같이 발달적으로 적절한 정의의 제시 문제는 ‘학교폭력’의 측정에서도 충분히 고려되어야 한다. 참고로, 청소년폭력예방재단(2014)의 실태조사 문항을 보면 학교폭력의 현 법률적 정의를 그대로 제시한 후 학생들의 학교폭력 관련 견해나 경험을 묻는 방식을 취하고 있는데 이러한 정의가 학생들 입장에서 이해가능하며, 정확한 보고를 촉진하는지의 여부는 후속 연구가 필요한 영역이다. 또한 괴롭힘이라는 용어에 대한 학생들의 부정적 정서 반응으로 인하여 괴롭힘 경험이 과소보고됨을 보여주는 해외

실증연구들(Kert et al., 2009; Vaillancourt et al., 2010)이 있는 바, 이러한 편향이 괴롭힘, 따돌림, 학교폭력 등 용어에서도 나타나는지 실증 연구를 통해서 밝혀질 필요가 있다.

넷째, 괴롭힘이라는 용어를 직접 사용하지 않는 심리척도를 통해 괴롭힘 경험을 측정하는 경우에도, 괴롭힘의 핵심 요소들을 구체적으로 조작화하는 전략을 통해 측정의 정밀성을 제고할 필요가 있다. 앞서 지적한 대로 괴롭힘의 핵심 요소를 적절히 조작화하지 못하는 경우, 괴롭힘과 다른 공격행동 간의 구분은 모호해 질 수 밖에 없으며, 이는 괴롭힘 측정의 타당도를 심각하게 저해하는 요인이 될 수 있다. 특히, 앞서 고찰하였듯이 힘의 불균형 요소의 적절한 조작화는 대표적인 괴롭힘 관련 자기보고식 척도들에서 거의 일관되게 무시돼온 요소(Swearer et al., 2010)로 괴롭힘의 개념화와 조작화 간의 큰 괴리를 보여주는 문제로 지적되고 있다(Furlong et al., 2010). 따라서 힘의 불균형이 괴롭힘을 다른 또래 공격행동과 구분 짓는 본질적 요소라면 현재 이 부분의 측정 전략에 있어서 분명한 개선이 요구된다. 가장 바람직한 방법은 이러한 요소를 반영하여 타당화한 자기보고식 괴롭힘 척도를 개발하는 것이지만, 대안적으로 기존의 괴롭힘 척도를 수정·보완하는 방법도 고려해 볼 수 있다. 예컨대, 괴롭힘 연구자는 힘의 불균형 요소를 묻는 부가적 차원의 질문(예: ‘상대방이 이러한 행동을 할 때 자신을 적절히 방어할 수 있었는가?’ 등)를 추가하여 매 괴롭힘 피해 경험 보고마다 가중치를 부여하는 방식의 채점을 통해 괴롭힘 측정의 정밀성을 제고할 수 있을 것이다.

의도성과 반복성의 경우도 조작화 측면에서 측정의 정밀성을 확보하기 위한 유사한 조치

들이 필요한 것으로 보인다. 예컨대, 피해학생의 의도성 귀인과 같은 인지적 과정이 괴롭힘 피해보고 과정에 영향을 미칠 수 있다는 사실(Camodeca et al., 2003)은 괴롭힘 유병률 추정이나 진단적 목적의 괴롭힘 평가(즉, 가해학생, 피해학생, 가해/피해학생 판정과 같은) 상황에서 이러한 가외변인을 고려할 필요성을 제기한다. 즉 의도성 귀인이 평가결과에 영향을 미치는 주요 조절변인이라면 괴롭힘 피해 경험 측정 시 피해학생의 귀인편향을 동시에 측정하는 평가 전략을 사용하여 측정결과의 해석에 활용한다면 결과해석의 정밀성을 높이는 데 도움이 될 수 있을 것이다. 또한 반복성의 조작적 정의를 연구자들 간에 합의한 후 괴롭힘 경험을 측정하는 조치도 필수적인 것으로 보인다. 현재로서는 먼 과거의 경험(예: ‘지난 1년 동안’)보다는 최근 과거의 경험(예: ‘지난 1달 동안’)을 묻는 것이 보다 정확한 보고를 유도하는지, 아니면 과잉보고를 유도하는지 명확하지 않으므로(Furlong et al., 2010; Morrison & Furlong, 2002), 외적 준거를 통해 평가하는 실증연구가 필요하겠지만 만일 최근의 시간참조틀이 보다 정확한 결과를 산출한다면 구체적인 절단점[가령, Solberg와 Olweus (2003)가 제시하는 ‘한 달에 2-3회’]에 대해서 연구자들 간에 먼저 합의하고 괴롭힘 실태조사의 문항을 이러한 시간참조틀에 맞추고 실시 간격도 이에 준하는 것이 보다 바람직할 것이다.

다섯째, Cornell과 Cole(2012)이 권고하였듯이 자기보고식 괴롭힘 측정 시 하위 유형의 특수성을 무시하고 단순히 다른 유형들과 합산하는 방식을 취하기보다는 이들 하위 유형에 대한 측정치를 별도로 제시하거나 아예 이들을 별개의 구성개념으로 취급하는 방법이 개념적

명료성 면에서나 실제적인 측면에서 더 나은 전략인 것으로 보인다. 이는 자기보고식 측정에서 직접적인 형태의 괴롭힘과는 달리 간접적 형태의 괴롭힘(관계적 괴롭힘 혹은 따돌림)이나 특수한 형태의 괴롭힘(사이버괴롭힘)은 구체적인 행동적 지표가 모호하거나(Sveinsson & Morris, 2007), 괴롭힘의 핵심적인 구성요소에 준하여 개념화하기 어렵다는 문제점(Smith, 2012)이 지적되고 있어서 이들 하위 유형을 괴롭힘에 포함시킬지의 여부에 대해 연구자들 간에 상충된 견해를 제시하고 있기 때문이다.

측정 전략적 측면

첫째, 설문조사방식의 자기보고식 평가에서 단일문항 접근과 다문항 접근은 각각 장단점을 가지고 있으나, 이 두 접근이 민감도와 특이도 면에서 차이가 있다는 일부 연구결과(Vaillancourt et al., 2010)가 제시되고 있으므로, 가능하다면 OBQ의 경우처럼 이 두 접근을 병행하는 것이 바람직할 것으로 보인다. 현재 국내의 청소년폭력예방재단(2014)나 교육부(2014)의 실태조사 질문지의 경우 전반적 단일질문(예: “학교폭력을 당한 경험이 있습니까?”)에 ‘없다’로 응답하는 경우, 이후 구체적인 하위유형에 대해서는 더 이상 응답하지 않도록 구성되어 있다. 따라서 단일 문항의 전반적 질문을 먼저 제시한 후 이러한 질문에 대한 동의 여부와 상관없이 추가로 구체적 피해/가해 유형별 경험을 묻는 다문항 질문에 답하도록 질문지를 재구성하여 피해 및 가해 경험을 측정해 볼 필요가 있다.

둘째, 익명성/비밀보장 여부, 실시자의 훈련, 실시 형태 등 자기보고식 평가도구를 실시하는 절차나 맥락과 관련된 요인들이 괴롭힘의

측정결과에 영향을 미치는 것으로 알려져 있으므로, 국내 괴롭힘 연구자들은 이러한 요인들에 보다 세심한 주의를 기울일 필요가 있다. 현재까지 보고된 연구 결과를 종합하면 학교 폭력 혹은 괴롭힘 실태조사에서 익명성의 보장이 기명 방식에 비해 괴롭힘 경험을 과잉 보고하도록 유도하는지 아니면 보다 정확한 보고를 유도하는지 명확하지 않다고 할 수 있다. 따라서 괴롭힘 연구자들은 Cornell과 Bandyopadhyay(2010)가 제안하였듯이, 익명성 실시에 대한 대안으로서 비밀보장 방식의 측정도 고려해 볼 필요가 있다. 무엇보다도 비밀보장 하에 괴롭힘을 측정함으로써 연구자는 학생의 사생활을 보호하면서 동시에 수집된 정보의 정확성을 입증 혹은 확인 가능하기 때문이다.

또한 괴롭힘이나 학교폭력 관련 설문조사 시 실시자에게 제공하는 비교적 간단한 훈련이 학생들의 자기보고 결과의 신뢰도를 높일 수 있다는 증거(Cross & Newman-Gonchar, 2004)가 제시되고 있는 바, 자기보고식 도구를 활용하는 괴롭힘 연구자들은 표준적인 조사 실시 프로토콜을 개발하여 이러한 기준을 준수하도록 조사실시자를 훈련시킴으로써, 그들이 자칫 지루하고 재미없는 조사에 학생들이 최선을 다해 신중하고 정확하게 반응하도록 고무시키는 역할을 할 수 있게 도와야 한다. 자기보고식 검사의 실시 형태(예: 온라인 실시 대 전통적 지필검사)에도 주목할 필요가 있다. 이러한 변인이 측정결과에 미치는 정확한 영향의 정도와 방향은 더 연구되어야 하는 부분이지만, 현재까지 보고된 연구결과들(Hilton et al., 2003; Turner et al., 1998)은 괴롭힘이나 학교폭력 실태조사를 가급적 하나의 방식으로 통일하여 실시하고 서로 다른 형태

로 실시한 실태조사의 결과를 직접 비교하는 것은 지양되어야 함을 의미한다. 참고로, 국내의 경우 교육부에서 실시하는 실태조사의 경우 2012년 2차 실시부터 모두 온라인 방식으로 전환되었고, 청소년폭력예방재단에서 실시하는 조사의 경우 온라인과 오프라인 조사를 병행하고 있지만 실시형태의 차이가 결과에 미친 영향에 대해서는 현재 알려져 있지 않다.

셋째, 자기보고식 괴롭힘 측정에서 무효반응, 민감성 증가, 요구특성 같은 응답자의 고유한 반응특성이 측정 결과의 신뢰도와 타당도를 위협하는 가외변인으로 밝혀짐에 따라 이러한 요인들을 제거하거나 통제할 수 있는 측정 전략을 마련할 필요가 있다. 예컨대, 무효반응을 걸러내는 선별기준을 조사의 착수 시점에 미리 마련하여 괴롭힘 유병률이 왜곡되어 추정되지 않도록 주의해야 할 것이다. 이러한 선별절차는 전수조사 같은 대규모 조사의 경우 자기보고식 도구 내에 무효반응을 걸러낼 수 있는 문항을 제작하여 함께 실시하는 방법을 고려해 볼 수 있을 것이며, 보다 작은 규모의 조사에서는 동일한 자기보고식 측정을 반복하거나 추후면접, 또래참조 평가 등을 추가하여 자기보고식 측정 결과의 신뢰성을 확인하는 절차를 마련할 수 있을 것이다. 또한 괴롭힘이나 학교폭력 예방 프로그램을 도입한 직후 해당 문제들에 대한 학생들의 민감성이 증가하고, 이후에 요구특성이 가외변인으로 작용하여 프로그램의 효과와 상관없이 일정한 패턴의 결과를 초래할 수 있다는 지적(Cornell & Cole, 2012)을 고려한다면, 예방 프로그램의 도입과 함께 실시하는 실태조사에서 얻은 결과를 해석하여 유병률의 증가나 감소에 대한 결론을 도출할 때 이러한 오염변인의 개입 가능성에 유념해야 할 것이다. 일례로,

국내의 한 실태조사결과(청소년폭력예방재단, 2014)를 분석해 보면 2006년 이후에 10% 내외의 피해율을 보이다가 2011년 말에 실시한 조사에서 18.3%로 급증한 결과를 보이다가 2012년과 2013년 조사에서는 각각 12.6%와 5.6%로 다시 급감하는 패턴을 보이고 있는데 이러한 결과를 해석할 때 고려할 점은 이러한 결과가 학교폭력의 실제적인 증감보다는 민감성과 요구특성을 반영하는 것일 가능성이 있다. 즉 이러한 패턴의 결과가 2011년 학교폭력 관련 자살 사건이 연쇄적으로 발생하면서 학교폭력에 대한 사회적 관심이 최고조에 달하고 동시에 각급 학교에서 학교폭력예방교육이 강화되면서 학생들의 민감성이 증가한 결과와 이후 점차로 요구특성에 따라 반응한 결과를 반영하였을 가능성을 배제할 수 없다는 점이다.

끝으로, 괴롭힘 연구자들은 자기보고식 도구를 통해 측정된 괴롭힘과 다른 자기보고식 측정치 간의 관계성을 연구할 때 공유방법분산의 문제에 대해 관심을 기울이고 괴롭힘 평가의 타당성을 제고하기 위해 측정방법을 다양화할 필요가 있다. 가령, 또래지명법(peer nomination)의 경우 자기보고식 측정과 유사한 측정 상의 문제를 안고 있고 또래끼리 서로 거명한다는 점에서 꺼려하는 측면이 있지만, 여러 명의 학생들로부터 누적된 자료에 기초하여 점수가 산정되므로 단일정보원(즉, 학생 자신)에 기초하는 방식보다는 측정 오차를 줄일 수 있고, 특히 괴롭힘 가해학생과 피해학생의 구체적 확인이 가능한 장점이 있다(Cornell & Cole, 2012). 따라서 괴롭힘 연구자는 익명으로 실시하는 자기보고식 평가에만 의존하기 보다는 비밀보장 방식의 자기보고와 또래지명법을 병용하여 잠재변인 수준에서 괴롭힘을 측정하고 다른 준거변인 간의 관계를 분

석함으로써 괴롭힘 연구에서 자기보고식 측정에만 의존함에 따르는 측정 상의 제한점을 개선하고 자기보고식 평가의 타당도를 제고하는 방안을 적극 고려할 필요가 있다.

결론

괴롭힘은 빈도, 의도, 힘의 역동 등 다면적 요소를 포함하는 특수한 형태의 공격행동으로 정확한 정의와 측정이 비교적 쉽지 않은 개념이다. 따라서 괴롭힘의 본질적 구성요소, 하위 유형 및 적절한 용어 등에 대해 다양한 견해가 존재해 왔으며, 특히 적절한 조작화의 문제와 평가 방법론적 문제는 현재 주요 쟁점이 되고 있다. 본 연구는 괴롭힘 개념화의 문제는 조작화 수준에서 해결하는 것이 적절하다고 전제하고, 괴롭힘이나 학교폭력 실태조사 및 각종 괴롭힘 관련 연구에서 가장 보편적으로 사용되고 있는 자기보고식 평가를 중심으로 괴롭힘 조작화의 문제와 측정 전략적 쟁점들을 관련 최근 연구문헌을 고찰함으로써 체계적으로 분석하였다. 본 연구는 괴롭힘의 자기보고식 측정에서 심리측정적 속성이 대체로 무시 혹은 경시되었으며, 이러한 도구들에서 나타나는 괴롭힘의 개념화와 조작화 간의 괴리 및 괴롭힘 측정 전략적 요인들은 괴롭힘 평가의 타당성을 저해하는 주요 요소임을 보여주었다. 따라서 자기보고식 평가의 정밀성을 높이고, 신뢰도 및 타당도 문제를 개선하는 노력이 없이는 괴롭힘의 유행률의 증가나 감소, 괴롭힘과 다른 변인간의 관계, 각종 교육 및 예방 프로그램의 효과 등을 알아보는 연구들은 과학적 엄밀성 면에서 심각한 제한을 받게 됨을 알 수 있다.

본 연구는 괴롭힘의 개념화 및 자기보고식 평가와 관련된 주요 쟁점을 종합적으로 검토하고 이러한 쟁점이 시사하는 개선방안을 찾아 구체적으로 제시하였다는 데에 의의가 있다. 특히 본 연구가 국내의 괴롭힘 연구자들이나 괴롭힘이나 학교폭력 관련 실태조사에 관여하는 교육현장의 연구책임자들에게 괴롭힘 평가의 과학적 엄밀성을 제고하는 기초 연구 자료로 활용되고 괴롭힘의 측정과 평가에 관한 국내 연구자들의 관심을 환기시키며, 아울러 괴롭힘 평가도구의 타당화 연구를 자극하기를 기대해 본다.

참고문헌

- 강진령, 유형근 (2000). 집단괴롭힘. 서울: 학지사.
- 교육부 (2014). 2014년 1차 학교폭력 실태조사 분석결과 발표 - 보도자료. <http://moe.go.kr>에서 2014년 8월 20일 자료 얻음.
- 권준모 (1999). 한국의 왕따 현상의 개념화와 왕따 연구의 방법론적 고찰. 한국심리학회지: 사회문제, 5(2), 59-72.
- 김종구, 박지현 (2014). 또래괴롭힘의 개념과 법적규율: 학교폭력예방 및 대책에 관한 법률과 관련하여. 일감법학, 27, 307-338.
- 김현주 (2003). 집단 따돌림에서의 동조집단 유형화 연구. 청소년복지연구, 5(2), 103-118.
- 김혜원 (2013). 청소년 학교폭력: 이해, 예방, 개입을 위한 지침서. 서울: 학지사.
- 네이버국어사전 (2014). <http://krdc.naver.com>에서 2014년 8월 20일 자료 얻음.
- 서미정 (2008). 방관자의 집단 특성에 따른 또래괴롭힘 참여 역할행동. 아동학회지,

- 29(5), 79-96.
- 송지윤 (2010). 중학생의 또래괴롭힘, 정서표현성, 사회적지지 및 신체증상경험의 관계. 이화여자대학교 교육대학원 석사학위청구논문.
- 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, 이정운, 유형근, 이은경, 고경희, 오혜영, 이유미, 김승혜, 최희영 (2014). 학교폭력 예방의 이론과 실제. 서울: 학지사.
- 이승연 (2013). 또래괴롭힘: 주변인 개입과 사회적 맥락 변화의 필요성. 한국심리학회지: 학교, 10(1), 59-81.
- 이춘재, 광금주 (1999). 학교에서의 집단 따돌림: 실태, 특성 및 대책. 서울: 집문당
- 이혜경, 김혜원 (2001). 초등학교생들의 집단괴롭힘 개해행동과 피해행동에 대한 사회적, 심리적 예측변인들: 학년과 성별을 중심으로. 한국심리학회지: 사회 및 성격, 15(1), 117-138.
- 정옥분, 정순화, 김경은, 박연정 (2008). 국내 집단따돌림 관련 연구의 경향분석. 인간발달연구, 15(1), 115-137.
- 정종진 (2013). 학교폭력의 개념. 송재홍, 김광수, 박성희, 안이환, 오익수, 은혁기, 정종진, 조봉환, 홍종관, 황매향 (공저), 학교폭력의 예방 및 대책(pp. 51-86). 서울: 학지사.
- 정태연, 김인경, 김은정 (2001). 집단 따돌림의 측정방법에 따른 행동평가의 차이: 자기평가와 동료평가의 비교. 한국심리학회지: 발달, 14(1), 145-149.
- 정한중 (2009). 학교폭력예방 및 대책에 관한 법률의 문제점과 개선방안. 경희법학, 44(1), 73-99.
- 조대훈, 김승혜, 강민수, 전연진, 이유미, 최희영 (2013). 학교 폭력 예방 및 대책. 서울: 박영사.
- 청소년폭력예방재단 (2014). 전국 학교폭력 실태 조사 연구. (재)푸른나무 청재단.
- 최보가, 임지영 (1999). 또래괴롭힘이 아동의 외로움과 불안에 미치는 영향. 대한가정학회지, 37(5), 111-121.
- 최은숙 (1999). 집단따돌림 가해, 피해 경향과 관련된 심리적 요인에 관한 일 연구. 서강대학교 교육대학원 석사학위논문.
- 최지영 (2008). 괴롭힘 상황에서의 아동의 가해 및 피해경험 측정방법 비교를 통한 심리적 특성 연구. 초등교육연구, 21(3), 339-358.
- Ahmad, Y., & Smith, P. K. (1990). Behavioral measures: Bullying in schools. *Newsletter of Association for Child Psychology and Psychiatry*, 12, 26-27.
- Arora, C. M. J. (1996). Defining bullying: Towards a clearer general understanding and more effective intervention strategies. *School Psychology International*, 17, 317-329.
- 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.

- Bovaird, J. A. (2010). Scales and surveys: some problems with measuring bullying behavior. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 277-292). New York, NY: Routledge.
- 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.
- Cook, C. R., Williams, K. R., Guerra, N. G., & Kim, T. E. (2010). Variability in the prevalence of bullying and victimization: A cross-national and methodological analysis. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 347-362). 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., & Loper, A. B. (1998). Assessment of violence and other high-risk behaviors with a school survey. *School Psychology Review, 27*, 317-330.
- 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.
- Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across seventy construct validation studies. *Journal of Marketing Research, 26*, 315-319.
- Crick, N. R., & Dodge, K. A. (1999). "Superiority" is in the eye of the beholder: A comment on Sutton, Smith, and Swettenham. *Social Development, 8*, 128-131.
- Crick, N. R., & Grotpeter, J. K. (1996). Children's treatment by peers: Victims of relational and overt aggression. *Development and Psychopathology, 8*, 367-380.
- Cross, J., & Newman-Gonchar, R. (2004). Data quality in student risk behavior surveys and administrator training. *Journal of School Violence, 3*, 89-108.
- Eliot, M., Cornell, D., Gregory, A., & Fan, X. (2010). Supportive school climate and student willingness to seek help for bullying and threats of violence. *Journal of School Psychology, 48*, 533-553.
- Espelage, D. L., & Swearer, S. M. (2003). Research on school bullying and victimization: What have we learned and where do we go from here? *School Psychology Review, 32*, 365-383.
- Frey, K. S., Hirschstein, M. K., Snell, J., Edstrom,

- L. V., Mackenzie, E., & Broderick, C. J. (2005). Reducing playground bullying and supporting beliefs: An experimental trial of steps to Step to Respect program. *Developmental Psychology, 41*, 479-491.
- 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.
- Griffin, R. S., & Gross, A. M. (2004). Childhood bullying: Current empirical findings and future directions for research. *Aggression and Violent Behavior, 9*, 379-400.
- 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.
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2003). Correspondence between self-reports of interpersonal violence. *Journal of Interpersonal Violence, 18*, 223-239.
- Jimerson, S. R., Nickerson, A. B., Mayer, M. J., & Furlong, M. J. (Eds.). (2012). *Handbook of school violence and school safety: International research and practice*. New York, NY: Routledge.
- Jimerson, S. R., Swearer, S. M., & Espelage, D. L. (Eds.). (2010). *Handbook of bullying in schools: An international perspective*. New York, NY: Routledge.
- Kert, A. S., Coddling, R. S., Tyrone, 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.
- Morrison, G., & Furlong, M. J. (2002, June). *Understanding the turing points in students' school discipline histories*. Paper presented at Safe Schools for the 21st Century, National Conference of the Hamilton Fish Institute, Monterey, California.
- 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. 9-33). New York,

- NY: Routledge.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.
- Reynolds, W. (2003). *Reynolds Bully Victimization Scales*. San Antonio, TX: The Psychological Corporation.
- Rigby, K. (2004). What it takes to stop bullying in schools: An examination of the rationale and effectiveness of school-based interventions. In M. J. Furlong, M. P. Bates, D. C. Smith, & P. Kingery (Eds.), *Appraisal and prediction of school violence* (pp. 165-191). Hauppauge, NY: Nova Science.
- Ross, D. M. (2003). *Childhood bullying, teasing, and violence: What school personnel, other professionals, and parents can do* (2nd ed.). Alexandria, VA: American Counseling Association.
- Schneider, K. T., Swan, S., & Fitzgerald, L. F. (1997). Job-related and psychological effects of sexual harassment in the workplace: Empirical evidence from two organizations. *Journal of Applied Psychology, 82*, 410-415.
- 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.
- Smith, D., & Ananiadou, K. (2003). The nature of school bullying and the effectiveness of school-based interventions. *Journal of Applied Psychoanalytic Studies, 5*, 189-209.
- Smith, P. K. (2012). Cyberbullying and cyber aggression. 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. 93-103). New York, NY: Routledge.
- 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.
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L.

- D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*, 867-873.
- 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.
- You, S., Furlong, M. J., Felix, E. D., Sharkey, J. D., Tanigawa, D., & Grief, J. (2008). Relations among school connectedness, hope, life satisfaction, and bullying victimization. *Psychology in the Schools, 45*, 446-460.
- Zins, J. E., Elias, M. J., & Mahler, C. A. (Eds.). (2007). *Bullying, victimization, and peer harassment: A handbook of prevention and intervention*. New York, NY: Routledge.
- 1차원고접수 : 2014. 09. 22.
수정원고접수 : 2014. 10. 21.
최종게재결정 : 2014. 10. 23.

Conceptualization and Self-Report Assessment of School Bullying: Current Issues and Future Directions

Donghyung Lee

Department of Education, Pusan National University

While the amount of research on bullying in schools has markedly increased in Korea and internationally, relatively little attention has been paid to its conceptualization and assessment practice. This study was to provide a systematic review on current issues in the conceptualization and self-report assessment of bullying with implications and future directions for research and practice. Recent domestic and international research publications were critically reviewed and three major issues (i.e., core components, subtypes and terminology) in the current conceptualization of bullying were identified and discussed, followed by self-report measurement/assessment issues in terms of the operationalization of core components and subtypes, single- vs. multiple-item approaches, contextual and procedural factors, response characteristics, and shared method variance. Implications and future directions for self-report assessment practice of bullying were derived and fully discussed especially for researchers and practitioners in Korea.

Key words : bullying, school violence, conceptualization, operationalization, self-report assessment