



2.1 Zielsetzung und Fragestellungen

Hauptzielsetzung der ersten Pilotstudie ist die Entwicklung und Erprobung eines Peer-Nominierungs-Fragebogens, mit dem Rollenverhalten im Bullying-Gefüge differenziert für verschiedene Formen von Bullying erfasst werden kann. Bislang vorliegende Instrumente fragen das Verhalten der Bystander von Bullying allgemein ab, ohne nach der Art des Bullying zu unterscheiden. Dies ist jedoch Voraussetzung dafür, das Rollenverhalten auch spezifisch für Cyberbullying untersuchen zu können. Dementsprechend wurde aufbauend auf dem Participant-Role-Ansatz (Salmivalli et al., 1996) ein Instrument entwickelt, welches Verhaltensweisen im Bullying-Gefüge getrennt nach Bullyingformen abfragt. In der Pilotstudie wurde zunächst eine Fassung für Schulbullying eingesetzt, um die generelle Praktikabilität des Erhebungsformates zu testen und den neuen Fragebogen zur Bestimmung der konvergenten Validität mit dem etablierten Participant Role Questionnaire (PRQ; Salmivalli et al., 1996) abzugleichen. Eine Erweiterung des Fragebogens um Items zu Cyberbullying wurde im Anschluss an die Pilotstudien in der Hauptstudie (siehe Kapitel 4) realisiert.

Konkrete Teilziele und Fragestellungen der Pilotierung des RoleGrid sind:

- (A) **Feststellen der Praktikabilität des Erhebungsformates:** Sind Definitionen, Instruktionen und Items für die angestrebte Zielgruppe verständlich?

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann
https://doi.org/10.1007/978-3-658-36619-3_2.

Kommen die Teilnehmenden mit dem Antwortformat der Peer-Nominierung zurecht?

- (B) **Bestimmen von Itemkennwerten, faktorieller Struktur und Testgütekriterien:** Welche deskriptiven Merkmale weisen die Items auf? Lässt sich die angenommene Faktorenstruktur entsprechend der fünf postulierten Rollen bestätigen? Wie hoch ist die interne Konsistenz der Rollen-Skalen?
- (C) **Festlegen eines geeigneten Kriteriums zur Identifikation der Rollen:** Welches Kriterium zur Zuordnung einer Rolle ist inhaltlich sinnvoll und pragmatisch umsetzbar? Wie hoch fallen die Prävalenzen der verschiedenen Rollen mit dem entsprechenden Kriterium aus?
- (D) **Ableich mit Participant Role Questionnaire und Selbstnominierungen:** Wie hoch sind die korrelativen Übereinstimmungen zwischen beiden Instrumenten? Wie hoch ist die Übereinstimmung hinsichtlich der Identifikation von Rollen? Inwiefern stimmen Selbst- und Fremdnominierungen überein?
- (E) **Analyse der Zusammenhänge zwischen Rollenverhalten und soziodemographischen sowie soziometrischen Merkmalen:** Zeigen sich die aus der Literatur zu erwartenden Zusammenhänge der Rollen mit Geschlecht, Alter, Zuneigung und Ablehnung?

Im Einzelnen wird angenommen, dass die Opfer von Bullying eher männlich sind, die Anzahl der Opfer in den höheren Klassenstufen geringer ist und die Opfer-Rolle mit geringer Akzeptanz und hoher Ablehnung assoziiert ist (vgl. Abschnitt 1.5.1). Zudem wird erwartet, dass Täter und Verstärker ebenfalls eher männlich sind und das Schikanieren in den mittleren Klassenstufen seinen Höhepunkt findet. Hinsichtlich des soziometrischen Status wird ein kontroverses Muster vermutet, bei dem die Täterschaft sowohl mit sozialer Akzeptanz als auch Ablehnung einhergeht, während der soziometrische Status der Verstärker gemäßigt und eher durchschnittlich ist (vgl. Abschnitt 1.5.2 und 1.5.3). Verteidigendes Verhalten sollte mit dem weiblichen Geschlecht, jüngerem Alter, hoher sozialer Akzeptanz und geringer Ablehnung assoziiert sein (vgl. Abschnitt 1.5.4). Weiterhin ist davon auszugehen, dass Außenstehende ebenfalls eher weiblich sind und dass passives Verhalten mit einem geringen sozialen Einfluss – also wenig Zuneigung aber auch wenig Ablehnung – einhergeht (vgl. Abschnitt 1.5.5). Als Hypothesen formuliert ergeben sich folgenden Annahmen für die Rollen im Bullying-Geschehen:

1.1. Die Opfer-Rolle ist mit dem männlichen Geschlecht assoziiert.

- 1.2. Die Opfer-Rolle ist negativ mit dem Alter assoziiert beziehungsweise in Klassenstufe 6 häufiger als in Klassenstufe 8.
- 1.3. Die Opfer-Rolle ist mit geringer Akzeptanz und hoher Ablehnung assoziiert.

- 2.1. Die Täter-Rolle ist mit dem männlichen Geschlecht assoziiert.
- 2.2. Die Täter-Rolle ist positiv mit dem Alter assoziiert beziehungsweise in Klassenstufe 8 häufiger als in Klassenstufe 6.
- 2.3. Die Täter-Rolle ist mit hoher Akzeptanz und hoher Ablehnung assoziiert.

- 3.1. Die Verstärker-Rolle ist mit dem männlichen Geschlecht assoziiert.
- 3.2. Die Verstärker-Rolle ist positiv mit dem Alter assoziiert beziehungsweise in Klassenstufe 8 häufiger als in Klassenstufe 6.
- 3.3. Die Verstärker-Rolle ist mit durchschnittlicher Akzeptanz und Ablehnung assoziiert.

- 4.1. Die Verteidiger-Rolle ist mit dem weiblichen Geschlecht assoziiert.
- 4.2. Die Verteidiger-Rolle ist negativ mit dem Alter assoziiert beziehungsweise in Klassenstufe 6 häufiger als in Klassenstufe 8.
- 4.3. Die Verteidiger-Rolle ist mit hoher Akzeptanz und geringer Ablehnung assoziiert.

- 5.1. Die Außenstehenden-Rolle ist mit dem weiblichen Geschlecht assoziiert.

- 5.2. Die Außenstehenden-Rolle soll explorativ hinsichtlich Altersunterschieden untersucht werden.
- 5.3. Die Außenstehenden-Rolle ist mit durchschnittlicher bis geringer Akzeptanz und Ablehnung assoziiert.

2.2 Konstruktion des Fragebogens

Der neu entwickelte Fragebogen – das RoleGrid – nutzt wie der PRQ (Salmivalli et al., 1996) die Einschätzungen der Klassenkameraden als Hauptinformationsquelle zur Identifikation der am Bullying-Geschehen beteiligten Rollen. Dies entspricht dem Grundgedanken des Participant-Role-Ansatzes, dass alle Klassenmitglieder in den Bullying-Prozess involviert oder sich dessen zumindest gewahr sind (Salmivalli et al., 1996, S. 2). Peer-Nominierungen oder Peer-Ratings zielen darauf ab, relativ stabile Merkmale wie typische Verhaltensmuster zu messen (Cronbach, 1970; Guilford, 1954; zitiert nach Olweus, 2010), was dem Konzept der Rolle als „Menge aufeinander bezogener, gelernter Verhaltensdispositionen, die Personen in interpersonale Beziehungen einbringen und aktualisieren“ (Frey & Greif, 1997, S. 253) gerecht wird. Vorteile gegenüber Selbstberichtsdaten sind zudem, dass die Gelegenheit zur positiven Selbstdarstellung umgangen wird und die Rollenzuordnung nicht auf der Auskunft von Einzelpersonen beruht (Badaly et al., 2013).

Um das Rollenverhalten möglichst effizient für verschiedene Formen von Bullying abfragen zu können, wurden grundlegende Änderungen vorgenommen. So werden im RoleGrid anstelle von Peer-Ratings *Peer-Nominierungen* eingesetzt, wie es für den PRQ von Schäfer und Korn (2004a) bereits erprobt und empfohlen wurde. Die Schülerinnen und Schüler werden also nicht angehalten jedes Klassenmitglied für jedes Item auf einer Rating-Skala einzustufen, sondern dürfen bei jedem Item eine beliebige Anzahl an Personen benennen. Dies ist weniger zeitaufwändig und für die Teilnehmenden mutmaßlich weniger ermüdend als die Rating-Methode. Zudem sehen sich die Heranwachsenden nicht gezwungen, auch zu Klassenmitgliedern ein Urteil abzugeben, die sie gar nicht gut kennen und einschätzen können. Die Entscheidung die Anzahl nennbarer Personen nicht zu beschränken beruht darauf, dass keine Vorannahmen über die Häufigkeit der Rollen getroffen werden sollten, um insbesondere zu vermeiden, dass bei den Befragten implizite Annahmen dazu entstehen, wie viele Personen es je Rolle geben müsste. Außerdem liegen Befunde vor, die bei unlimitierten verglichen

mit limitierten Nominierungen eine höhere Reliabilität, günstigere Verteilungseigenschaften und eine bessere ökologische Validität nachweisen (Bukowski, Cillessen & Velásquez, 2012; Gommans & Cillessen, 2015; Marks, Babcock, Cillessen & Crick, 2013).

Des Weiteren wird auf die Differenzierung zwischen Tätern und Assistenten (Salmivalli et al., 1996) verzichtet und diese beiden Rollen als (*Mit-*)*Täter zusammengefasst*. Dass die Trennung von Tätern, Assistenten und Verstärkern in der Literatur nicht unumstritten ist, wurde bereits in Abschnitt 1.3.5 angemerkt. Exploratorische Faktorenanalysen deuten wiederholt auf eine Vier-Faktoren-Lösung, bei der Täter, Assistenten und Verstärker auf einem gemeinsamen Pro-Bullying-Faktor laden (Goossens et al., 2006; Knauf et al., 2017; Salmivalli et al., 1998; Schäfer & Korn, 2004a; Sutton & Smith, 1999). Konfirmatorische Analysen untermauern teils die Zusammenfassung der Pro-Bullying-Rollen (Crapanzano, Frick, Childs & Terranova, 2011; Tani et al., 2003), teils eine Unterscheidung von Tätern und Mitläufern (Goossens et al., 2006) und bei separater Betrachtung der Pro-Bullying-Items auch eine Ausdifferenzierbarkeit in Täter, Assistenten und Verstärker (Schäfer & Korn, 2004a). Methodische Arbeiten, die mit allen Items des Instruments mehrere Modelle systematisch vergleichen, liegen nicht vor. Bei genauerer inhaltlicher Betrachtung der Items ist eine fehlende Abgrenzung zwischen den Pro-Bullying-Rollen nicht verwunderlich. Wird danach gefragt, wer beim Bullying mitmacht und dem Täter hilft (Assistent) oder wer zuschaut und lacht (Verstärker), ist davon auszugehen, dass dies auch auf diejenige zutrifft, die sonst als anführende Täter in Erscheinung treten. Im RoleGrid wird dementsprechend bei der Frage nach den Verstärkern darauf geachtet zu betonen, dass diese selbst nicht aktiv beim Bullying mitmachen. Der Versuch zwischen aktiven Tätern und Mittätern einerseits und passiven Verstärkern andererseits zu unterscheiden, obgleich diese Unterteilung in der Literatur bislang nicht vorgenommen wurde, hat konzeptionelle sowie pragmatische Gründe. Zum einen ist die Erfassung von Bullying-förderndem Bystander-Verhalten zentral für die Fragestellung dieser Arbeit, zum anderen ist es plausibel, dass den Befragten die Unterscheidung zwischen aktivem und passivem Verhalten leichter fällt als beispielsweise die Unterscheidung zwischen Anführern und Mitläufern.

Anders als im PRQ werden im RoleGrid die Rollen nicht mit mehreren Items zu verschiedenen Verhaltensweisen (z. B. „wer ist dabei, um zuzusehen?“ und „wer lacht?“ für die Verstärker-Rolle) abgefragt, sondern mit einer einzelnen möglichst prägnant formulierten Frage pro Rolle („Wer schaut dabei zu oder lacht darüber, macht aber selbst nicht aktiv mit?“). Die Verwendung von Einzelitems in Peer-Nominierungs-Verfahren ist nicht unüblich (Badaly et al., 2015; Hartl, Laursen & Cillessen, 2015), da es sich dabei aus messtheoretischer Perspektive

um eine Aggregation über mehrere Nominierende hinweg – statt über mehrere Items– handelt (Marks et al., 2013). Nichtsdestoweniger ist es für die interne Konsistenz ratsam, mehrere Items einzusetzen, um mehr Daten zu sammeln (Marks et al., 2013). Dieser Empfehlung kommt das RoleGrid nach, indem jede Rolle für *verschiedene Formen von Bullying* abgefragt wird. Für Schulbullying sind es in Anlehnung an Tabelle 1.1:

- direkte körperliche Angriffe auf die Person: Körperverletzung (physisches Bullying)
- direkte verbale oder non-verbale Angriffe: Verletzung durch sprachlichen, mimischen oder gestischen Ausdruck (verbales Bullying)
- indirekt-verbale Angriffe auf den Status: Rufschädigung (relationales Bullying)
- Verwehren von Teilhabemöglichkeiten: Ausgrenzung (isolierendes Bullying)
- Angriffe auf das Eigentum: Eigentumsverletzung (materielles Bullying).

Dies entspricht – bis auf die Angriffe auf das Eigentum – auch den Formen, die im Revised Olweus Bullying Questionnaire zur Definition herangezogen werden (Olweus, 2010). Die Ergänzung um Eigentumsverletzungen wurde bereits in einer Reihe von Studien vorgenommen (Knauf et al., 2017; Kristensen & Smith, 2003; Oldenburg et al., 2015; Smith & Shu, 2000; Verlinden et al., 2014). Neu ist, dass im RoleGrid bei der Frage nach direkt-verbalem Bullying auch gestische Beleidigungen, Erniedrigungen oder Bedrohungen explizit mitberücksichtigt werden. Der Einfachheit halber wird im Folgenden vorwiegend von verbalen Angriffen gesprochen, obgleich verbale und nonverbale Aktionen gemeint sind. Eine inhaltliche Unterscheidung verschiedener verbaler Bullyingformen, wie sie verschiedentlich vorgenommen wurde (bspw. rassistisches Bullying bei Smith & Shu, 2000), erfolgte nicht. Für Cyberbullying ist mit Ausnahme der physischen Angriffe auf Person und Eigentum eine ähnliche Unterteilung möglich, da die verbalen und nonverbalen Formen auch über digitale Medien ausgeübt werden können: direkte verbale Attacks, Rufschädigung sowie Ausgrenzung. Diese sind in der Pilot-Version des RoleGrid noch nicht enthalten, da das RoleGrid zunächst mit dem PRQ abgeglichen werden sollte, welcher sich ausschließlich auf Schulbullying bezieht. Grundsätzlich entspricht die einzelne Abfrage verschiedener Formen von Bullying, jeweils mit Beispielen versehen, auch den Empfehlungen der Centers for Disease Control and Prevention (Gladden et al., 2014). Kowalski et al. (2014) argumentieren, dass die Nutzung mehrerer Items zu verschiedenen Bullyingformen anstelle eines globalen Items die Sensitivität und Reliabilität des Instruments begünstigen.

Die Items sind rasterförmig in einer Matrix aus fünf Bullyingformen nach Art der Schädigung (in Zeilen: Körperverletzung, verbale Verletzung, Rufschädigung, Ausgrenzung, Eigentumsverletzung) und fünf Rollen (in Spalten: Opfer, (Mit-) Täter, Verstärker, Verteidiger und Außenstehende) angeordnet, woraus sich die Bezeichnung als RoleGrid (Rollenraster) ableitet. Die erste Spalte nimmt dabei immer die Frage nach Opfern derartiger Schikanen ein, um abzuklären, ob diese Art von Bullying in der Klasse überhaupt vorkommt („Wer aus deiner Klasse wird schikaniert, indem...“). Außerdem soll so gewährleistet werden, dass die Befragten bei der Einschätzung der übrigen Rollen immer das konkrete Opfer sowie die Art der Angriffe im Bewusstsein haben und nicht auf implizite Persönlichkeitstheorien zurückgreifen. Dies soll verhindern, dass Personen mit allgemein dissozialen, prosozialen oder passiven Verhaltenstendenzen benannt werden, die das entsprechende Verhalten jedoch in der konkreten Bullying-Situation gar nicht zeigen. Des Weiteren soll mit dieser Art der Strukturierung eine eindeutigere Zuordnung zu den Rollen erreicht und Doppelrollen reduziert werden, welche bei bisherigen Verfahren nicht unüblich sind (Knauf et al., 2017; Schäfer & Korn, 2004a).

Um sicherzustellen, dass sich die Antworten der Befragten tatsächlich auf Bullying und nicht auf aggressives Verhalten im Allgemeinen beziehen, wird dem Fragebogen eine Definition vorangestellt, wobei wie bei Schäfer und Korn (2004a) die deutsche Übersetzung von Bullying (Schikanieren) verwendet wird. Der oft auch alltagssprachlich verwendete Begriff Mobbing wird nicht genannt, da nicht davon ausgegangen werden kann, dass Jugendliche den Begriff entsprechend wissenschaftlicher Auffassung verwenden (Guerin & Hennessy, 2002). Die Vermeidung des bekannten Begriffs soll einer Verzerrung durch Vorwissen vorbeugen. Die vorgegebene *Definition von Schikanieren* deckt sowohl den Wiederholungscharakter als auch das Machtungleichgewicht ab. Zusätzlich werden verhaltensnahe Beispiele (vgl. behavioral descriptors; Cook, Williams, Guerra & Kim, 2010) genannt, um mögliche Äußerungsformen des aggressiven Verhaltens zu spezifizieren. Das Zeitkriterium wird nachfolgend spezifiziert als mindestens wöchentlich über mehrere Wochen und entspricht damit den Empfehlungen von Olweus (2010). Die komplette Pilot-Fassung des RoleGrid ist in Anhang A1 im elektronischen Zusatzmaterial (Electronic Supplementary Material = ESM) einsehbar.

2.3 Methode

Zur Pilotierung des RoleGrid in der Fassung für Schulbullying (RoleGrid-off) wurde der Fragebogen im Juli sowie im Oktober und November 2017 in einer Stichprobe von Schülerinnen und Schülern der sechsten und achten Klassenstufe eingesetzt. Die Erhebungen fanden innerhalb einer Unterrichtsstunde statt und wurden von fünf angehenden Lehrkräften im Rahmen ihrer wissenschaftlichen Arbeiten zur Ersten Staatsprüfung durchgeführt. Alle Arbeiten wurden von der Autorin betreut. Die benötigten Unterlagen wie der Fragebogen, Schulanschriften und Elternbriefe wurden von der Autorin entwickelt und den Erhebungspersonen zur Verfügung gestellt. Zudem wurden die Erhebungspersonen von der Autorin zum Thema Bullying und zum Einsatz des Fragebogens intensiv geschult. Schulen in Baden-Württemberg wurden von den Lehramtsstudierenden persönlich, telefonisch oder per E-Mail kontaktiert, um die Teilnahmebereitschaft zu erfragen. Allen Schulen wurde ein Informationsschreiben zukommen gelassen, in welchem Inhalt und Zweck der Befragung sowie das Vorgehen erklärt wurden. Auch ethische Aspekte wie Freiwilligkeit, Anonymität und Datenschutz wurden angesprochen. War die Teilnahmebereitschaft seitens der Schule gegeben, wurde über Elternbriefe das Einverständnis der Erziehungsberechtigten eingeholt. Lag die Einverständniserklärung bei weniger als der Hälfte der Klassenmitglieder vor, wurde die Befragung nicht durchgeführt, da geringe Teilnahmequoten bei Peer-Nominierungen mit geringen Reliabilitäten einhergehen (Marks et al., 2013). Dies war bei einer Schulklasse der Fall. Die bei der Befragung anwesenden Schülerinnen und Schüler bekamen nach der Befragung als Dankeschön eine kleine Süßigkeit. Es wurden Anlaufstellen für den Fall von Bullying genannt, die sowohl von selbst Betroffenen aber auch von Klassenkameraden genutzt werden können. Die Klassenlehrkräfte erhielten ein kleines Informationsschreiben zum Thema Bullying und nach Abschluss der wissenschaftlichen Arbeiten eine kurze Rückmeldung zu den Ergebnissen der jeweiligen Abschlussarbeit.

2.3.1 Stichprobe

Insgesamt wurden 23 Klassen von fünf verschiedenen Schulen erhoben. Es nahmen drei Realschulen, eine Werkrealschule und eine Gemeinschaftsschule teil. Die teilnehmenden Klassen umfassten 576 Schülerinnen und Schüler (47.6 % weiblich), von denen 489 (84.9 %) an der Befragung teilnahmen. Gründe für die

Nicht-Teilnahme waren Abwesenheit in der Erhebungsstunde (10.2 %), die fehlende Einverständniserklärung Erziehungsberechtigten (3.6 %) oder die eigene Entscheidung des Jugendlichen, nicht teilnehmen zu wollen (1.2 %). Die Teilnahmequoten verschiedener Klassen schwankten zwischen 62 % und 100 %. Die Klassengröße lag bei 18 bis 29 Schülerinnen und Schülern ($M = 25.0$; $SD = 3.79$). Die minimale Teilnehmerzahl pro Klasse betrug 14 Personen. Es wurden 12 sechste Klassen ($n = 302$) und 11 achte Klassen ($n = 274$) erhoben. Das Alter der befragten Schülerinnen und Schüler lag zwischen 10 und 16 Jahren ($M = 12.5$; $SD = 1.24$). Informationen zur Herkunft (Geburtsland der Eltern) lagen bei 476 Jugendlichen vor, von denen 22.3 % einen Migrationshintergrund (mindestens ein Elternteil nicht in Deutschland geboren) hatten. Häufigste außerdeutsche Geburtsländer der Eltern waren Russland, Kasachstan und Polen.

2.3.2 Instrumente und Datenerhebung

Nachdem in Abschnitt 2.2 die Entwicklung des RoleGrid detailliert beschrieben und begründet wurde, sollen an dieser Stelle das letztendlich in der Pilotierung eingesetzte Instrument sowie weitere verwendete Instrumente und Items dargestellt werden (siehe Anhang A1 im ESM). Der Fragebogen beginnt mit einer allgemeinen Beschreibung, worum es in der Befragung geht, gefolgt von einer ausführlichen Definition des Begriffs Schikanieren in folgendem Wortlaut:

Schikanieren ist, wenn einem Schüler oder einer Schülerin über einen *längeren Zeitraum* durch Mitschüler/innen *immer wieder* körperlicher oder psychischer Schaden zugefügt wird und er oder sie sich nicht dagegen wehren kann.

Es folgen Erläuterungen und Beispiele zu den verschiedenen Arten, wie diese Schädigung erfolgen kann. Der Fokus liegt dabei auf dem Opfer („Jemand kann in der Schule beispielsweise dadurch schikaniert werden, dass...“), um die Systematik der Angriffe hervorzuheben. Zudem wird deutlich gemacht, dass es sich nicht um Schikanieren handelt, wenn Schüler/innen streiten, die gleich stark sind. Die einführenden Informationen sind auf der ersten Seite des Fragebogens abgedruckt und werden zusätzlich mündlich von den Erhebungsleitern mit den Klassen durchgesprochen. Anschließend werden mit den Teilnehmenden die Aspekte der Freiwilligkeit, Ehrlichkeit und Anonymität geklärt. Schülerinnen und Schüler dürfen selbst ankreuzen, ob sie an der Befragung teilnehmen möchten oder nicht. Als erstes werden daraufhin *soziodemografische Angaben* zu Geschlecht, Alter und

Herkunft abgefragt. Zwei *soziometrische Items* erfragen, mit welchen Klassenkameraden der Schüler oder die Schülerin am liebsten (Liked-Most-Nominierungen) oder am wenigsten gern (Liked-Least-Nominierungen) etwas zusammen macht. Nach diesem allgemeinen Teil folgen die beiden Fragebögen zur Erfassung der Rollen: das neu entwickelte RoleGrid sowie zur Validierung eine leicht modifizierte Version des PRQ (Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a).

Die pilotierte Version des *RoleGrid* fragt mit insgesamt 25 Items rasterförmig zu fünf Formen von Schulbullying (Körperverletzung, verbale Verletzung, Rufschädigung, Ausgrenzung, Eigentumsverletzung) fünf Rollen (Opfer, (Mit-)Täter, Verstärker, Verteidiger und Außenstehende) ab. Opfer-Rolle und Bully-Rolle werden dabei mit den Fragen „Wer aus deiner Klasse wird schikaniert, indem...?“ und „Von wem?“ abgedeckt. Die Bystander-Rollen werden prägnant mit jeweils einer Formulierung umschrieben (Verstärker: „Wer schaut dabei zu oder lacht darüber, macht aber selbst nicht aktiv mit?“; Verteidiger: „Wer tröstet das Opfer oder versucht das Schikanierten zu stoppen?“; Außenstehende: „Wer bekommt davon nichts mit oder hält sich raus und unternimmt gar nichts?“). Die Teilnehmenden werden instruiert, diejenigen Klassenkameraden zu nominieren, auf welche das beschriebene Verhalten passt. Zusätzlich ist in einem offenen Antwortfeld die Möglichkeit gegeben zu vermerken, was die übrigen Klassenmitglieder währenddessen machen.

Der *Participant Role Questionnaire* (PRQ; Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a) verfährt ebenfalls nach dem Prinzip der Peer-Nominierung. Die konstituierenden Items einer Rollenskala differenzieren jedoch nur bei Täter und Opfer nach Art der Schikanen. Die Bystander-Skalen setzen sich aus Items zu den verschiedenen für eine Rolle charakteristischen Verhaltensweisen zusammen. Insgesamt umfasst der PRQ 22 Items, die sich zu sechs Skalen gruppieren lassen: Opfer-Skala (5 Items, z. B. „Über wen werden hinter seinem/ihrer Rücken immer wieder Gerüchte erzählt?“), Täter-Skala (5 Items, z. B. „Wer schikaniert andere, indem er/sie tritt, rumschubst oder stößt?“), Assistenten-Skala (3 Items, z. B. „Wer hilft dem Täter oder der Täterin?“), Verstärker-Skala (3 Items, z. B. „Wer lacht, wenn jemand schikaniert wird?“), Verteidiger-Skala (3 Items, z. B. „Wer setzt sich ein, dass die anderen mit dem Schikanierten aufhören?“) und Außenstehenden-Skala (3 Items, z. B. „Wer hält sich immer aus der Situation raus, wenn jemand schikaniert wird?“).

RoleGrid und PRQ wurden zur Vermeidung von Reihenfolge-Effekten klassenweise ausbalanciert: in 11 Klassen wurde zuerst das RoleGrid ($n = 197$) und in 12 Klassen ($n = 291$) zuerst der PRQ bearbeitet. Um bei den Peer-Nominierungs-Verfahren die Anonymität der Teilnehmenden zu gewährleisten,

wurden Codierlisten zur Hilfe genommen, in denen jedem Klassenmitglied eine Codenummer zugeordnet war. In den Fragebögen durften keine Namen, sondern ausschließlich die Codenummern eingetragen werden. Diese wurden nach Abschluss der Erhebung vernichtet, wodurch Rückschlüsse auf Einzelpersonen ausgeschlossen sind. Für die Befragung wurde gegebenenfalls die Sitzordnung angepasst und wenn möglich Sichtschutz zwischen den Teilnehmenden aufgestellt, um eine ungestörte Bearbeitung des Fragebogens zu ermöglichen.

Die Analyse von Itemkennwerten, Faktorenstruktur und Reliabilitäten des RoleGrid sowie dessen Auswertungsmöglichkeiten sind Hauptgegenstand dieser Pilotstudie. Aus diesem Grund werden diesbezügliches Vorgehen und die Ergebnisse ausführlich in den nachfolgenden Kapiteln behandelt. Da die Skalenstruktur des PRQ nicht unumstritten ist (Crapanzano et al., 2011; Goossens et al., 2006; Sutton & Smith, 1999), wird auch diese erneut geprüft und die Reliabilitäten der Skalen erst im Ergebnisteil berichtet.

2.3.3 Auswertungsstrategie und statistische Analysen

Alle statistischen Analysen wurden mit der Statistiksoftware IBM SPSS Statistics 25 sowie der Softwareumgebung R Version 3.6.0 (2019-04-26) durchgeführt. Die Daten der Pilotstudie zum RoleGrid wurden zunächst deskriptiv ausgewertet, wobei die Itemverteilungen sowie Itemkennwerte betrachtet wurden. Die Intraklassenkorrelationen und Designeffekte für Item- und Skalenwerte konnten übersichtlich mit einer Funktion von Michael Hock (persönliche Kommunikation) generiert werden. Zur Ermittlung der faktoriellen Struktur wurden mit Hilfe von SPSS exploratorische Faktorenanalysen und mit dem R-Paket lavaan Version 0.6-5 (Rosseel et al., 2019; siehe auch Rosseel, 2012) konfirmatorische Faktorenanalysen gerechnet. Zur Berücksichtigung der genesteten Datenstruktur wurde auf die Prozedur von F. L. Huang (2017) zurückgegriffen. Zur graphischen Darstellung der Faktorenstruktur wurde das Paket semPlot (Epskamp, 2019) genutzt. Für die Punktschätzer der internen Konsistenzen (Cronbachs Alpha, McDonalds Omega) der Skalen sowie deren Konfidenzintervalle (bias-corrected and accelerated Bootstrapping) wurde das MBESS-Paket (Kelley, 2019) hinzugezogen.

Zur Einschätzung der Prävalenz der verschiedenen Rollen wurden zwei Kriterien zur Rollenzuordnung einander gegenübergestellt. Zum einen erfolgte die Zuordnung nach dem etablierten Vorgehen der Standardisierung (z-Kriterium; Salmivalli et al., 1996; Schäfer & Korn, 2004a), bei dem die Rollenzuweisung

stattfindet, sobald der standardisierte Skalenwert über dem Durchschnitt liegt – das heißt größer als null ist – und auf keiner anderen Skala ein ähnlich hoher Wert erzielt wird (Differenz von mindestens 0.1). Zum anderen wurde eine Zuordnung auf Grundlage eines absoluten Cut-off-Wertes von drei Nennungen, die für mindestens eine Form von Bullying erreicht werden müssen, vorgenommen (min3-Kriterium). Verschiedene Vorgehensmöglichkeiten bei der Rollenzuordnung werden ausführlich in Abschnitt 2.3.4 reflektiert. Basierend auf den beiden hier gewählten Zuordnungsstrategien wurde die kategoriale Übereinstimmung zwischen RoleGrid und PRQ (Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a) überprüft. Zusätzlich wurde als Index für die konvergente Validität auch die korrelative Übereinstimmung zwischen den Rollenskalen von RoleGrid und PRQ (Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a) herangezogen. Zur Beurteilung der Kriteriumsvalidität wurden weiterhin regressionsanalytisch die Zusammenhänge der RoleGrid-Skalen mit soziodemografischen und soziometrischen Variablen bestimmt. Hierzu wurde die `glmer`-Funktion des `lme4`-Paketes Version 1.1-21 (Bates et al., 2019) in R genutzt (siehe auch Bates, Maechler, Bolker & Walker, 2015).

2.3.4 Exkurs: Reflexion verschiedener Kriterien für die Rollenzuordnung

Das von den Autorinnen und Autoren des PRQ (Salmivalli et al., 1996) ursprünglich vorgeschlagene und am häufigsten eingesetzte Kriterium zur Identifikation der Rollen im Bullying-Geschehen sieht vor, dass ein Klassenmitglied überdurchschnittliche Werte auf einer Skala erzielen muss, um die entsprechende Rolle zugeordnet zu bekommen. Praktisch bedeutet dies, dass alle Skalenwerte klassenweise z-standardisiert werden, sodass die Klassenmittelwerte gleich null sind, und die Rollenzuordnung dann erfolgt, wenn der Schüler oder die Schülerin einen Skalenwert von über null erreicht (Salmivalli et al., 1996; Schäfer & Korn, 2004a). Ist der Score mehrerer Skalen überdurchschnittlich, wird die Rolle mit dem höheren Wert zugeordnet sofern die Differenz mindestens 0.1 beträgt. Sind zwei Skalenwerte überdurchschnittlich und die Differenz kleiner 0.1, kann die Person als nicht klar zuzuordnen betrachtet werden (Salmivalli et al., 1996) oder gegebenenfalls eine Zweitrolle vergeben werden (Schäfer & Korn, 2004a).

Diese auf Standardisierung beruhende Vorgehensweise ist jedoch nicht unumstritten. Bereits Sutton und Smith (1999) merkten kritisch an, dass die Methode anfällig für Verzerrungen durch Extremwerte sei. Einzelne extrem hohe Werte führen dazu, dass auch der Klassenschnitt höher ausfällt und andere Kinder

mit hohen, aber eben nicht überdurchschnittlichen Werten somit nicht für eine Rolle identifiziert werden können. Umgekehrt ist auch denkbar, dass in Klassen mit geringem Durchschnitt Heranwachsende mit absolut betrachtet wenig stark ausgeprägtem Verhalten (z. B. nur ein oder zwei Nennungen) für eine Rolle identifiziert werden (vgl. Olweus, 2010; Pouwels et al., 2016). Zudem werden intraindividuelle Unterschiede zwischen den verschiedenen Skalen bei einer Standardisierung nicht erhalten. Beispielsweise könnte eine Person mit weniger Opfern als Täter-Nominierungen als Opfer identifiziert werden, wenn der Klassenschnitt für Viktimisierung deutlich unter dem der Täterschaft liegt. Goossens et al. (2006) weisen darauf hin, dass man mit der Standardisierungsmethode in jeder Klasse Vertreter aller Rollen findet und dieses Kriterium zu einer ähnlichen Rollenverteilung in jeder Klasse führt. Dies hat zum einen den Nachteil, dass Unterschiede zwischen verschiedenen Klassen nivelliert werden (Olweus, 2010) und zeitlichen Veränderungen nicht angemessen abgebildet werden können (Goossens et al., 2006). Zum anderen kann an dieser Stelle auch Salmivallis Grundannahme, dass in jeder Klasse alle Rollen vertreten sind, hinterfragt werden (Pouwels et al., 2016). Atria und Kolleginnen (2007) fanden hinsichtlich Täterschaft und Viktimisierung eine hohe Variabilität zwischen verschiedenen Schulklassen und zweifeln infolgedessen an, dass die Participant Roles in jeder Klasse gefunden werden könnten.

Als Alternative erprobten Sutton und Smith (1999) eine Standardisierung an der Gesamtstichprobe oder eine Zuordnung ohne Standardisierung basierend auf dem intraindividuell höchsten Skalenscore. Bei Caravita et al. (2019) erfolgt eine Rollenzuordnung, wenn der Rollenscore mindestens eine Standardabweichung über dem Durchschnitt der Altersgruppe liegt. Da diese Methoden grundlegende Mängel des ursprünglichen Vorgehens nicht ausräumen können (z. B. Abhängigkeit des individuellen Scores von der übrigen Stichprobe bei jeglicher Form der Standardisierung) beziehungsweise neue Probleme mit sich bringen (z. B. fehlende interindividuelle Vergleichbarkeit bei intraindividuellem Kriterium), haben sie sich in der weiteren Forschung nicht durchgesetzt.

Mehrfach vorgeschlagen und eingesetzt wurde hingegen ein prozentuales Zuordnungskriterium (Goossens et al., 2006; Pouwels et al., 2016): Die Anzahl erhaltener Nominierungen wird durch die Anzahl der Befragten abzüglich des Nominierten ($n-1$) geteilt. Eine Rolle wird dann zugeordnet, wenn die Person von einem Mindestanteil der Klassenmitglieder (z. B. mindestens von 10 %) nominiert wurde. Ein solches Vorgehen wurde mit einer Hürde von 30 % von Salmivalli et al. (1996) zur Identifikation der Opfer genutzt. Ohne Nachteile ist allerdings auch diese Methode nicht. Zwar wird es gerade als Zweck dieser Methode erachtet, mögliche Effekte der Klassengröße auszugleichen (Velásquez, Bukowski &

Saldarriaga, 2013), doch bedeutet dies unweigerlich auch, dass die Stimme eines Befragten je nach Klassengröße unterschiedlich zählt. Die Nennung, die ein Schüler abgibt, fällt in kleinen Klassen stärker ins Gewicht als in großen. Während in einer kleinen Klasse mit 15 Schülern fünf Nennungen ausreichen, um das 30 %-Kriterium zu erfüllen, müssen es in einer großen Klasse mit 30 Schülern immerhin neun Nennungen sein. Unter der Grundannahme des Participant-Role-Ansatzes, dass nahezu jeder von den Schikanen mitbekommen, scheint dies ein sinnvolles Vorgehen. Bei Einbezug verdeckter Formen von Bullying – gerade im Kontext von Cyberbullying – muss die Annahme jedoch hinterfragt werden. Auch wenn es um die Identifikation der Bystander-Rollen geht, muss davon ausgegangen werden, dass Bystander-Verhalten nicht immer öffentlich gezeigt wird und nicht jedes Klassenmitglied über alle übrigen Bescheid weiß. Gerade in großen Klassen kann es also sein, dass sich die Klassenmitglieder untereinander weniger kennen und daher seltener nominieren (vgl. Velásquez et al., 2013). Obgleich es in größeren Klassen also potenziell mehr Personen gibt, die eine Person X nennen könnten, gibt es gleichzeitig auch mehr Personen, die genannt werden können.

Eine weitere Möglichkeit der Rollenzuordnung, die jedoch bislang wenig Anwendung findet, basiert auf einem absoluten Cut-off-Wert. Es wird also eine Mindestzahl an Nominierungen festgelegt, die für eine Rollenzuweisung notwendig ist. Die offensichtliche Schwierigkeit bei dieser Methode liegt in der Entscheidung für einen geeigneten absoluten Trennwert. Schuster (1999) erachtet Schülerinnen oder Schüler ab fünf Nennungen als Opfer, Obermann (2011b) hingegen weist die Täter- oder Opfer-Rolle bereits ab zwei Nennungen zu. Branson und Cornell (2009) vergleichen verschiedene Cut-off-Werte und kommen zu dem Schluss, dass ein Trennwert von zwei Nennungen die Unterschiede zwischen Opfern und Nicht-Opfern beziehungsweise Tätern und Nicht-Tätern hinsichtlich externer Validierungskriterien (z. B. aggressive Einstellungen, Schulverweise, Notenschnitt, Depression) maximiert. In einer Validierungsstudie von Phillips und Cornell (2012) führten erfahrene Beratungslehrer (school counselors) Gespräche mit nominierten Schülerinnen und Schülern, um den Verdacht auf Viktimisierung zu prüfen. Von denjenigen mit zwei Nominierungen bestätigte sich der Verdacht in weniger als der Hälfte der Fälle. Je höher die Anzahl an Peer-Nominierungen, desto höher war die Rate derer, die sich im Gespräch als Opfer herausstellten. Die Gruppe der bestätigten Opfer hatte durchschnittlich 5.5 Nennungen erhalten, die Gruppe der unbestätigten Opfer nur 2.9. Zusammenfassend ist die Forschungslage zur Validität und Eignung verschiedener absoluter Cut-off-Werte äußerst dünn – einheitliche Empfehlungen liegen nicht vor. Bei der Wahl eines geeigneten Trennwertes für Forschungszwecke geht es letztlich auch darum, die Waage zu halten zwischen Sensitivität (möglichst viele Personen einer bestimmten Rolle

als solche zu erkennen) und Spezifität (möglichst wenigen Personen fälschlich eine Rolle zuzuschreiben). Aus der Logik heraus, dass es sich bei den Participant Roles um soziale Rollen handelt, die von der Gruppe relativ konsistent wahrgenommen werden sollten, und in Einklang mit der Annahme, dass nahezu die ganze Gruppe über das Geschehen Bescheid weiß, wurde in der Publikation von Knauf, Eschenbeck und Käser (2017) das vergleichsweise strenge Kriterium von mindestens sechs Nennungen pro Verhaltensindikator gewählt. Hervorzuheben ist, dass dieses Mindestkriterium sich nicht auf den Skalenmittelwert, sondern auf die einzelnen Items bezieht. Die Rolle wurde also unabhängig davon zugeordnet, wie viele Facetten des Rollenverhaltens gezeigt wurden. Insgesamt konnten mit diesem Kriterium 57.5 % der Stichprobe eine Rolle zugeordnet werden (verglichen mit 87 % bei Salmivalli et al., 1996). Insbesondere die wenig saliente Gruppe der Außenstehenden fiel mit 8.8 % klein aus (verglichen mit 23.7 % bei Salmivalli et al., 1996). Diese Ergebnisse werfen die Frage auf, ob die Einbußen an Sensitivität gerade für unauffälliges Bystander-Verhalten durch ein dermaßen hoch angesetztes Kriterium nicht zu groß sind. Verfolgt man eine andere Logik, in der es weniger darum geht, dass die Rollen von einem Großteil der Klasse konsistent wahrgenommen werden müssen, sondern eher darum, die Verlässlichkeit der Nominierung zu gewährleisten, lässt sich auch ein Kriterium von mindestens drei Nennungen rechtfertigen. Möchte man sich für die Rollenzuordnung nicht auf die Aussage einzelner Schülerinnen und Schüler verlassen, sollte diese Einschätzung von mindestens zwei weiteren Klassenkameraden, die in diesem Sinne als Zeugen betrachtet werden können, bestätigt werden.

2.4 Ergebnisse

Das erste wichtige Ziel der Pilotstudie war die Prüfung, ob der neue Fragebogen praktikabel für den Einsatz an weiterführenden Schulen ist. Die Erhebungen verliefen reibungslos und ohne größere Zwischenfälle. Aus eigenen Pretestungen und Gesprächen mit den Erhebungspersonen ergibt sich das Bild, dass viele Schülerinnen und Schüler dem Thema aufgeschlossen und interessiert begegnen, es aber durchaus auch Heranwachsende gibt, die sich lustlos oder gelangweilt geben. Insbesondere in den sechsten Klassen schien das Ausfüllen beider Fragebögen für einige Befragten anstrengend und hat etwas länger gedauert als in den achten Klassen. In manchen Klassen war eine etwas intensivere Unterstützung in Form von wiederholten oder ausführlicheren Erläuterungen notwendig. Dies konnte gut gewährleistet werden, da die meisten Erhebungen in Zweiertteams durchgeführt wurden. Es traten keine gravierenden Schwierigkeiten im Umgang

mit dem Instrument auf. Mit dem Erhebungsformat der Peer-Nominierung kamen die Teilnehmenden zurecht. Der Begriff des Schikanierens sowie das Prinzip der Peer-Nominierung wurden verstanden. In der abschließenden Frage, welchen Fragebogen die Schülerinnen und Schüler für die nächsten Untersuchungen empfehlen würden, bevorzugte ein Großteil (52 %) den PRQ, wohingegen nur 27 % eine klare Empfehlung für das RoleGrid gaben. Die übrigen hatten keine Präferenz (kreuzten keine oder beide Antwortoptionen an). Berücksichtigt man die Reihenfolge der Fragebogenbearbeitung zeigt sich, dass die Vorliebe für den PRQ besonders hoch ausfällt, wenn dieser als erstes bearbeitet wurde (62 %). Wurde der PRQ als zweites ausgefüllt, lag die Bevorzugung des PRQ bei nur noch 36 % und 30 % gaben dem RoleGrid den Vorzug.

2.4.1 Itemstatistiken des RoleGrid und des PRQ

Die Verteilungen der Items wurden mittels Itemstatistiken (Mittelwert, Streuung, Schiefe und Exzess), Tests auf Normalverteilung sowie visuell mit Hilfe von Histogrammen, QQ-Plots und Boxplots inspiziert. Alle Items des RoleGrid und auch des PRQ (Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a) weisen eine linkssteile, schmalgipflige Verteilung auf, die signifikant von der Normalverteilung abweicht (Kolmogorov-Smirnov-Test und Shapiro-Wilk-Test). Modus ist durchgängig der Wert null. Im elektronischen Zusatzmaterial finden sich zwei umfassende Tabellen mit allen relevanten Itemstatistiken für RoleGrid und PRQ (Tabelle B1 und Tabelle B2).

Im Hinblick auf mögliche absolute Kriterien für die Rollenzuordnung ist es sinnvoll die Häufigkeitsverteilungen näher zu betrachten. Das von Knauf et al. (2017) eingesetzte Kriterium von mindestens sechs Nennungen wird für das RoleGrid je nach Item nur von 0–4 % der Schülerinnen und Schüler erfüllt. Bei 16 von 25 Items wird dieses absolute Kriterium von weniger als einem Prozent der Stichprobe erfüllt. Dies betrifft vor allem die Bystander-Rollen. Bei den Items für die Verstärker-Rolle bei relationalem Bullying (Gerüchten) und für die Außenstehenden bei ausgrenzendem und materiellem Bullying wird keine Person mehr als fünfmal genannt. Höhere Raten erreichen nur die Opfer- und die Bully-Rolle. Ein weniger extremes Bild ergibt sich für den PRQ (Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a). Hier werden für den Großteil der Items (20 von 22) mindestens 1 % der Stichprobe mehr als fünfmal genannt. Auch hier sind es Items für Verstärker und Außenstehende, für die weniger Personen genannt werden. Die Rate an Personen, die das Kriterium von mindestens sechs Nennungen erfüllen, ist bei der Frage danach, wer mitmacht (Assistenten-Rolle)

mit knapp 7 % am höchsten. Setzt man das in Abschnitt 2.3.4 ebenfalls diskutierte Kriterium von mindestens drei Nominierungen an, liegen die Raten je nach Item des RoleGrid zwischen 0.5 % und 15.8 %. Bei nur zwei Items (Verstärker bei ausgrenzendem Bullying und Außenstehende bei materiellem Bullying) werden weniger als 1 % der Stichprobe mindestens dreimal nominiert. Beim PRQ erfüllen je nach Item zwischen 3.8 % (Außenstehenden-Item: schaut weg) und 21.2 % (Verteidiger-Item: tröstet) das Kriterium von mindestens drei Nominierungen.

Da die Erhebungen in Schulklassen stattfanden, wurde anhand von Intraklassenkorrelationen (ICC) und Designeffekten geprüft, in welchem Ausmaß dadurch die Annahme der Unabhängigkeit der Messungen verletzt wurde (vgl. Maas & Hox, 2005). Dazu wurde das R-Paket multilevel Version 2.6 (Bliese, 2016) genutzt. Die Mehrheit der Items von sowohl RoleGrid als auch PRQ weisen große Designeffekte (> 2) auf. Für die Opfer-Items beider Instrumente zeigen sich hingegen durchweg kleinere Designeffekte. Einige der Opfer-Items weisen sogar negative ICC auf, was bedeutet, dass hinsichtlich der Viktimisierung die Varianz innerhalb der Schulklassen größer ist als zwischen den Schulklassen (vgl. Bliese, 2000).

Im Wesentlichen ergeben sich somit drei Herausforderungen für die Datenanalysen: Erstens sind die Variablen nicht normalverteilt, weshalb auch nicht von einer für gängige Methoden der Faktorenanalyse notwendigen multivariaten Normalverteilung ausgegangen werden kann (Bühner, 2006; Tabachnick & Fidell, 2014). Die Itemverteilungen zeichnen sich insbesondere durch eine enorme Zahl an Null-Werten einerseits und einige extrem hohe Werte andererseits aus. Dies führt zum zweiten Punkt, dass es sich zwar um intervallskalierte, nicht aber um stetige Daten handelt. Nicht-lineare Transformationen zur Normalisierung, wie sie im Falle nicht-normalverteilter Daten oft empfohlen werden (Bühner, 2006; Kline, 2011; Tabachnick & Fidell, 2014), hätten somit eine Herabsetzung des Skalenniveaus zur Folge sowie den Verlust des natürlichen Nullpunkts und der natürlichen Einheit der Skala (Anzahl der Nennungen). Daten mit Absolutskalenniveau wie Häufigkeiten sollten keiner Transformation unterzogen werden (Bühner, 2006). O'Hara und Kotze (2010) warnen zudem davor diskrete Count-Daten einer Log- oder Quadratwurzel-Transformation zu unterziehen, da diese insbesondere bei geringem Mittelwert zu deutlich verzerrten Schätzungen führen, wenn die Daten aus einer negativen Binomialverteilung stammen. Dies wirft die Frage nach der anzunehmenden zugrundeliegenden Verteilung der hier untersuchten Merkmale in der Population auf, welche im Rahmen dieser Arbeit nicht abschließend geklärt werden kann. Doch ist anzunehmen, dass auch die Populationsverteilung nicht normal ist. Drittens liegt eine genestete Datenstruktur mit

erhöhten Designeffekten vor. Die Unabhängigkeit der Messungen ist somit verletzt. Wie im Einzelnen mit den Besonderheiten in der Datenverteilung und der genesteten Datenstruktur umgegangen wird, wird im jeweiligen Abschnitt mit direktem Bezug auf das jeweilige Verfahren geschildert.

2.4.2 Faktorenstruktur des RoleGrid

Ogleich konkrete Annahmen zur faktoriellen Struktur des RoleGrid bestehen, sollte mittels exploratorischer Faktorenanalyse ein erster Eindruck von der Datenstruktur gewonnen werden. Da nach der Betrachtung der univariaten Verteilungen, nicht von einer multivariaten Normalverteilung ausgegangen werden kann (Kline, 2011), welche Voraussetzung für die Maximum-Likelihood-Methode ist, wurde die Methode der Hauptachsenanalyse gewählt, die laut Bühner (2006) weder Normalverteilung noch Intervallskalenniveau voraussetzt und auch bei einheitlich schief verteilten Items akzeptable Ergebnisse hervorbringt. Zwar erlaubt dieses Verfahren ohne Verteilungsannahmen keine statistische Prüfung der Modellparameter und der Modellpassung (Schmitt, 2011), genügt jedoch dem Zweck eine Idee von Faktorenzahl und Itemzuordnung zu bekommen.

Das nötige Ausmaß linearer Zusammenhänge zwischen den Variablen ist laut Kaiser-Meyer-Olkin-Koeffizient ($KMO = .837$), Measure of Sample Adequacy (MSA aller Items $\geq .77$) und Bartlettstest auf Sphärizität ($\chi^2_{(300)} = 8054.5$; $p < .001$) gegeben. Die anfänglichen Kommunalitäten der Items liegen zwischen $h^2 = .368$ und $h^2 = .783$. Kaiser-Guttman-Kriterium, Scree-Plot, Parallelanalyse und MAP-Test (O'Connor, 2000) deuten auf eine Fünf-Faktoren-Lösung. Nach obliquen Rotation (Promax, $Kappa = 4$) zeigen diese eine klare Fünf-Faktoren-Struktur, bei der jeweils die fünf erwarteten Items auf einem Rollenfaktor laden. Eine Tabelle mit den Ladungen und Kommunalitäten aller Items findet sich im elektronischen Zusatzmaterial (Tabelle B3). Die Faktoren klären gemeinsam 59.4 % der Gesamtvarianz auf. Bei Reproduktion der Korrelationsmatrix liegen 6 % nicht-redundante Residuen vor.

Um abzuschätzen inwiefern Zusammenhänge aufgrund der Varianz zwischen den Klassen, also aufgrund der genesteten Datenstruktur, oder durch die Extremwerte überschätzt wurden, wurden zwei weitere exploratorische Hauptachsenanalysen durchgeführt. Die eine Analyse wurde mit klassenweise zentrierten Items durchgeführt, wodurch Mittelwertunterschiede zwischen den Klassen eliminiert wurden. Für die andere Analyse wurden alle Werte, die größer als drei waren, auf den Wert drei gesetzt. Dieses Vorgehen gleicht dem Winsorisieren

(Leonhart, 2017), jedoch sollte das untere Ende der Verteilung unverändert bleiben. Damit die Items die gleiche Spannweite erhalten und somit vom Format eher einer vierstufigen Rating-Skala gleichen, wurden die Daten nicht etwa um einen bestimmten Prozentsatz winsorisiert, sondern der absolute Wert drei gewählt, der schon als Zuordnungskriterium für die Rollen diskutiert wurde (Abschnitt 2.3.4) und welcher in den Boxplots der meisten Items bereits als Ausreißer gilt. Dies kommt auch einem Vorschlag von Tabachnick und Fidell (2014) zum Umgang mit Extremwerten gleich. Beide Vorgehensweisen – klassenweise Zentrierung sowie das Trimmen – führen zu nahezu identischen Ergebnissen wie die Analyse der Rohwerte. Auch hier zeigt sich die erwartete Fünf-Faktoren-Struktur. Die aufgeklärte Varianz ist etwas geringer (58.8 % bei Gruppenzentrierung, 51.9 % bei einseitiger Winsorisierung) und der Prozentsatz nicht-redundanter Residuen etwas höher (9 % bei Gruppenzentrierung, 10 % bei einseitiger Winsorisierung). Insgesamt erweist sich die Fünf-Faktoren-Lösung somit als stabil und scheint nicht wesentlich durch die genestete Datenstruktur oder die Extremwerte verzerrt zu sein.

Die konfirmatorische Faktorenanalyse mit Maximum-Likelihood-Schätzer und robusten Standardfehlern sowie Santorra-Bentler-skaliertes Teststatistik ($\text{estimator} = \text{„MLM“}$) zeigt eine mäßige Passung der Fünf-Faktorenstruktur: $\chi^2_{(265)} = 654.99$, $p < .001$ (Skalierungsfaktor = 1.1), robuster CFI = .940, robuster TLI = .932, robuster RMSEA = .053, 90 % CI [.048, .058] und SRMR = .041. Es liegt kein exakter Modell-Fit vor – der χ^2 -Test zeigt signifikante Abweichungen von der Datenstruktur an und das Vertrauensintervall des RMSEA schließt nicht den Wert null ein (Bühner, 2006; Schermelleh-Engel, Moosbrugger & Müller, 2003). Da der χ^2 -Signifikanztest jedoch stark abhängig von Modellkomplexität und Stichprobengröße ist, wird verschiedentlich empfohlen, zusätzlich auch die Fit-Indizes zu betrachten (Bühner, 2006; Schermelleh-Engel et al., 2003). Eine Übersicht zu den Fit-Indizes und deren Interpretation in Anlehnung an Schermelleh-Engel et al. (2003) gibt Tabelle 2.1. Das Verhältnis zwischen χ^2 -Teststatistik und Freiheitsgraden kann als deskriptiver Goodness-of-Fit-Index interpretiert werden und deutet in diesem Fall auf eine akzeptable Passung (Schermelleh-Engel et al., 2003). Die inkrementellen Fit-Indizes (CFI, TLI) liegen etwas unter dem akzeptablen Bereichs, die Verbesserung gegenüber einem restriktiven Nullmodell ist also nicht zufriedenstellend. Die absoluten Fit-Indizes (RMSEA, SRMR), welche Aufschluss über die Passung im Vergleich mit einem saturierten Modell geben, liegen hingegen im akzeptablen bis guten Bereich.

Tabelle 2.1 Interpretation der Fit-Indizes in Anlehnung an Schermelleh-Engel et al. (2003)

| Fit-Index | akzeptable Passung | gute Passung |
|-------------------------------------------------|--------------------|--------------|
| χ^2 | $\leq 3df$ | $\leq 2df$ |
| Comparative Fit Index (CFI) | $\geq .95$ | $\geq .97$ |
| Tucker Lewis Index (TLI) | $\geq .95$ | $\geq .97$ |
| Root Mean Square Error of Approximation (RMSEA) | $\leq .08$ | $\leq .05$ |
| Standardized Root Mean Residual (SRMR) | $\leq .10$ | $\leq .05$ |

Die Parameter-Schätzungen für das Fünf-Faktoren-Modell sind in Abbildung C1 des Anhangs im elektronischen Zusatzmaterial dargestellt. Die höchste Faktorladung weist jeweils das Item zu verbalem Bullying auf. Der Opferfaktor korreliert negativ mit Verstärker- und Verteidiger-Faktor, der Bullyfaktor korreliert positiv mit dem Verstärker-Faktor und negativ mit Verteidiger- und Außenstehenden-Faktor und untereinander korrelieren Verstärker- und Verteidiger-Faktor negativ (alle $p < .005$). Stärkste Korrelation war jene zwischen Bully- und Verstärkerfaktor.

Zur Berücksichtigung der Mehrebenenstruktur der Daten kann die von F. L. Huang (2017) bereitgestellte Prozedur für das R-Paket lavaan genutzt werden, welche die Analysen auf Grundlage der gepoolten Within-Kovarianzmatrix berechnet, jedoch nur mit dem normalen ML-Schätzer. Alternativ kann die Analyse analog zum exploratorischen Vorgehen auch mit den gruppenzentrierten Werten erfolgen. Beide Vorgehensweisen führen zur gleichen Einschätzung der Modellpassung. Abweichungen ergeben sich bei den Fit-Indizes erst ab der dritten Nachkommastelle. Auch das beschriebene Muster an Faktorladungen und Interfaktorkorrelationen bleibt bestehen. Beide Alternativstrategien kommen zu identischen standardisierten Ladungs- und Korrelationsschätzungen.

2.4.3 Faktorenstruktur des PRQ

Da die Skalenstruktur des PRQ nicht unumstritten ist (Crapanzano et al., 2011; Goossens et al., 2006; Sutton & Smith, 1999), wurden auch für diesen Fragebogen exploratorische und konfirmatorische Faktorenanalysen gerechnet. Ausreichende lineare Zusammenhänge zwischen den Variablen sind laut Kaiser-Meyer-Olkin-Koeffizient ($KMO = .901$), Measure of Sample Adequacy (MSA aller Items $\geq .70$) und Bartlettstest auf Sphärizität ($\chi^2_{(231)} = 9440.8$; $p < .001$) gegeben. Kaiser-Guttman-Kriterium, Scree-Plot, Parallelanalyse und MAP-Test

(O'Connor, 2000) implizieren eine Vier-Faktoren-Lösung. Die Hauptkomponentenanalyse mit obliquer Rotation (Promax, Kappa = 4) ergibt eine klare Struktur für vier Rollen: Pro-Bullying-Akteure, Opfer, Verteidiger und Außenstehende. Diese vier Faktoren klären gemeinsam 65.1 % der Varianz auf. Eine Gruppenzentrierung oder das Trimmen der Werte ändern an Faktorenstruktur und Itemzuordnung nichts, allein die aufgeklärte Varianz ist bei getrimmten Werten um fast 10 Prozentpunkte geringer.

Die konfirmatorische Faktorenanalyse mit MLM-Schätzer zeigt eine mäßige Passung der Vier-Faktorenstruktur: $\chi^2_{(203)} = 576.71$, $p < .001$ (Skalierungsfaktor = 1.5), robuster CFI = .936, robuster TLI = .927, robuster RMSEA = .069, 90 % CI [.062, .076] und SRMR = .053. Bei der Schätzung des Sechs-Faktoren-Modells hingegen zeigen sich Schätzprobleme (die Kovarianzmatrix der latenten Variablen ist nicht positiv definit), welche vermutlich auf eine hohe Interkorrelation der Pro-Bullying-Faktoren zurückgehen ($r = .94-.99$). Die Ladungen und Kommunalitäten der exploratorischen Faktorenanalyse sowie die Parameterschätzungen der konfirmatorischen Faktorenanalyse sind für die Vier-Faktoren-Lösung im elektronischen Zusatzmaterial (Tabelle B4) einsehbar.

2.4.4 Reliabilitäten und Skalenstatistiken des RoleGrid und des PRQ

Die internen Konsistenzen sowie weitere Skalenstatistiken für alle Skalen des RoleGrid und des PRQ sind in Tabelle 2.2 aufgeführt. Mit einem Cronbachs Alpha zwischen .83 und .90 sind die internen Konsistenzen für das RoleGrid durchweg als gut zu beurteilen (Kline, 2011). Auch beim PRQ liegen die Werte von Cronbachs Alpha im akzeptablen bis hervorragenden Bereich (Kline, 2011). Daran ändert sich (mit Ausnahme der Außenstehenden-Skala des PRQ) nichts, wenn der Reliabilitätsschätzung anstelle von Itemrohwerten die klassenweise zentrierten Peer-Nominierungen oder die auf maximal drei Nennungen getrimmten Werte zugrunde gelegt werden (vergleiche Vorgehen bei exploratorischer Faktorenanalyse in Abschnitt 2.4.2 und 2.4.3). Die nähere Betrachtung der Bedeutsamkeit einzelner Items für die Skalen zeigt, dass das Entfernen eines Items in der Regel zu einer Abnahme und nie zu einer substanziellen Verbesserung der internen Konsistenz führt. Das trennschärfste Item für die Skalen des RoleGrid ist jeweils jenes zu verbalem Bullying. Da von einer tau-kongenerischen Messung ausgegangen werden muss (unterschiedliche Faktorladungen/Trennschärfen der Items), stellt Cronbachs Alpha eine Mindestschätzung der Reliabilität dar

Tabelle 2.2 Interne Konsistenzen und Skalenzustatistiken des RoleGrid und des PRQ (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a)

| Skala (Itemzahl) | Cronbachs Alpha [CI] | McDonalds Omega [CI] | M (SD) | Min | Max | Schiefe | Exzess | ICC | Design |
|----------------------|-------------------------|-------------------------|-------------|-----|------|---------|--------|-------|--------|
| RoleGrid | | | | | | | | | |
| Opfer (5) | .90 [.84; .94] | .91 [.86; .95] | 0.45 (1.30) | 0.0 | 17.8 | 6.66 | 64.76 | -.017 | 0.59 |
| Bully (5) | .89 [.85; .92] | .90 [.86; .94] | 0.83 (1.38) | 0.0 | 15.2 | 4.82 | 34.92 | .046 | 2.12 |
| Verstärker (5) | .83 [.77; .89] | .84 [.78; .91] | 0.42 (0.65) | 0.0 | 6.8 | 3.66 | 23.79 | .125 | 4.01 |
| Verteidiger (5) | .87 [.83; .91] | .88 [.84; .92] | 0.43 (0.75) | 0.0 | 7.4 | 3.39 | 18.58 | .092 | 3.22 |
| Außenstehende (5) | .85 [.81; .88] | .87 [.86; .90] | 0.58 (0.72) | 0.0 | 5.8 | 2.38 | 8.32 | .244 | 6.88 |
| PRQ | | | | | | | | | |
| Opfer (5) | .89 [.85; .93] | .92 [.88; .95] | 0.69 (1.60) | 0.0 | 18.8 | 5.55 | 41.61 | .005 | 1.12 |
| Pro-Bullying (11) | .94 [.93; .96] | .95 [.94; .96] | 0.96 (1.42) | 0.0 | 10.2 | 3.06 | 12.21 | .067 | 2.60 |
| Verteidiger (3) | .90 [.88; .92] | .91 [.89; .93] | 1.31 (1.69) | 0.0 | 11.3 | 2.49 | 8.19 | .070 | 2.67 |
| Außenstehende (3) | .73 [.68; .79] | .79 [.73; .84] | 1.15 (1.15) | 0.0 | 7.7 | 1.89 | 4.80 | .173 | 5.17 |

Anmerkung. In eckigen Klammern angeführt sind die mittels bias-corrected and accelerated (BC_a) Bootstrapping ermittelten 95 %-Konfidenzintervalle. Skalenscores wurden als Mittelwert der jeweiligen Peer-Nominierungen berechnet; ICC = Intraklassenkorrelation; Design = Designeffekt; N = 576.

(Bühner, 2006, S. 133). Eine angemessene Alternative zur Schätzung der internen Konsistenz bietet der Koeffizient Omega, welcher von einem kongenerischen Messmodell ausgeht und ein geringeres Risiko birgt, die Reliabilität zu unter- oder überschätzen (Dunn, Baguley & Brunson, 2014). In Tabelle 2.2 berichtet wird McDonalds Omega, dessen Formel die konservativste Berechnungsmethode für Omega darstellt (Jorgensen, Pornprasertmanit, Schoemann & Rosseel, 2019). Ebenfalls in Tabelle 2.2 angeführt sind die Intraklassenkorrelationen und Designeffekte für die Skalen. Diese sind mit Ausnahme der Opfer-Skala deutlich erhöht. Insbesondere bei der Außenstehenden-Skala geht ein beachtlicher Anteil der Varianz auf die Klassenzugehörigkeit zurück.

Die Korrelationen zwischen den Skalenwerten des RoleGrid sind in Tabelle B5 des elektronischen Zusatzmaterials einsehbar. Neben Pearsons Korrelationskoeffizient ist auch Kendalls Rangkorrelationskoeffizient aufgeführt, welcher besser für nicht-normalverteilte Daten geeignet ist. Ebenfalls betrachtet wurden die Korrelationen für die Mittelwertskalen der getrimmten Daten, um den Einfluss von Extremwerten abschätzen zu können. Pearsons Korrelationen bewegen sich mit getrimmten und ungetrimmten Skalen in einer ähnlichen Größenordnung und zeigen ein fast identisches Signifikanzmuster, die Werte für Kendalls Tau sind mit getrimmten und ungetrimmten Skalen nahezu identisch¹. Die Befunde zu den getrimmten Skalen werden daher nicht extra in der Tabelle aufgeführt. Wiederum berichtet werden beide Korrelationskoeffizienten für die klassenweise zentrierten Mittelwertskalen. Dies trägt der genesteten Datenstruktur Rechnung, indem Niveauunterschiede zwischen den Klassen eliminiert und somit die Zusammenhänge innerhalb der Klassen (within-group correlations) wiedergegeben werden. Dabei treten – ungeachtet des Korrelationstyps – die Kontraste zwischen den Rollen stärker hervor: Mit Ausnahme der positiven Korrelation zwischen Täter- und Verstärker-Skala sind fast alle Skalen negativ korreliert. Liegt ein Schüler mit einer Skala über dem Klassenschnitt, liegt er oder sie in den anderen Skalen eher unter dem Klassenschnitt. Mit nicht-zentrierten Skalenwerten sind weniger negative Korrelationen zwischen den Skalen signifikant. Die Rangkorrelationen zeigen generell ein etwas anderes Muster signifikanter Korrelationen als Pearsons Korrelation, meist jedoch nur, wenn es um geringe Korrelationen geht (Signifikanzniveau von $p < .05$ wird beim einen Korrelationstyp erreicht und beim

¹ Durch das Trimmen der Daten geht das Intervallskalenniveau verloren, weshalb Rangkorrelationen angemessener sind als Pearsons Korrelation. Bei der Berechnung von Kendalls Tau werden die Daten als Ränge behandelt und nur die Anzahl der Inversionen und Proversionen betrachtet, was schon bei Nutzung der Rohwertskalen den Einfluss von Extremwerten reduziert, sodass das zusätzliche Trimmen keinen weiteren Effekt hat.

anderen nicht). Die Rangkorrelation zwischen Täter- und Opfer-Skala wird allerdings auf einem Niveau von $p < .001$ signifikant, obgleich kein signifikanter linearer Zusammenhang (Pearsons Korrelation) zu bestehen scheint.

Unabhängig von Korrelationstyp und Zentrierung zeigt sich eine positive Korrelation zwischen Täter- und Verstärker-Skala und die Verteidiger-Skala korreliert negativ mit Verstärker- sowie Täter-Skala. Dies sind auch die höchsten Interfaktorkorrelationen (vgl. Abschnitt 2.4.2 und Abbildung C1, ESM). Ganz ähnliche Muster lassen sich für die Skalen des PRQ erkennen (siehe Tabelle B6, ESM), mit dem Unterschied, dass hier recht konsistent auch Verteidiger- und Außenstehenden-Skala sowie Opfer- und Außenstehenden-Skala positiv korrelieren.

2.4.5 Prävalenz der Rollen laut RoleGrid und PRQ

Da kein Goldstandard zur Bestimmung der Rollen auf Basis der Skalenscores existiert, werden im Folgenden die Prävalenzen nach zwei verschiedenen Vorgehensweisen präsentiert. Für die Vergleichbarkeit mit bisherigen Forschungsbefunden wird die weitverbreitete Methode der klassenweisen z-Standardisierung eingesetzt (z-Kriterium), bei der eine Rollenzuweisung erfolgt, wenn ein Skalenwert höher als der Klassendurchschnitt ist und auf keiner anderen Skala ein ähnlich hoher Wert erzielt wurde (Salmivalli et al., 1996). Als Alternative erfolgt eine Rollenzuordnung auf Basis eines absoluten Kriteriums: Eine Rolle wird dann zugewiesen, wenn die Person von mindestens drei Klassenkameraden für mindestens einen Rollenindikator genannt wurde (min3-Kriterium).

Bei beiden Zuordnungskriterien besteht die Möglichkeit, dass einer Person keine Rolle zugeordnet wird – wenn kein Skalenwert über dem Schnitt der Klasse liegt oder für kein Item mindestens drei Nominierungen vorliegen. Auch kann es sein, dass die Rollenzuordnung nicht eindeutig ist, weil für mehrere Skalen ein ähnlich hoher überdurchschnittlicher Wert erreicht wurde oder weil für mehrere Items unterschiedlicher Skalen mindestens drei Nennungen gegeben wurden. Die Häufigkeiten der Rollen gemäß dieser beiden Zuordnungskriterien sind in Tabelle 2.3 dargestellt.

Tabelle 2.3 Prävalenzen der Rollen in Prozent laut RoleGrid und PRQ. (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a) nach zwei verschiedenen Zuordnungskriterien

| Rolle laut RoleGrid | Zuordnungskriterium | | Rollen laut PRQ | Zuordnungskriterium | |
|----------------------------|----------------------------|---------|------------------------|----------------------------|---------|
| | mind. 3 | $z > 0$ | | mind. 3 | $z > 0$ |
| Opfer | 6.4 | 9.7 | Opfer | 3.5 | 12.3 |
| Bully | 16.0 | 13.9 | Pro-Bullying | 21.7 | 24.0 |
| Verstärker | 4.0 | 17.0 | | | |
| Verteidiger | 7.8 | 20.5 | Verteidiger | 11.1 | 22.4 |
| Außenstehende | 6.6 | 22.9 | Außenstehende | 9.0 | 22.9 |
| uneindeutig | 15.1 | 3.3 | uneindeutig | 29.5 | 2.8 |
| keine Rolle | 44.1 | 12.7 | keine Rolle | 25.2 | 15.6 |

Anmerkung. Rollenzuordnung erfolgte entweder, wenn der klassenweise standardisierte Skalenwert größer als null war ($z > 0$) oder wenn für mindestens einen Rollenindikator mindestens drei Nennungen vorlagen (mind. 3); $N = 576$.

Klassenweise betrachtet bedeuten diese Prävalenzen, dass es – je nach Instrument und Kriterium – in jeder Klasse durchschnittlich ein bis drei Opfer, fünf bis acht Pro-Bullying-Akteure, zwei bis sechs Verteidiger sowie zwei bis sechs Außenstehende gibt. Die genauen Werte für die beiden Instrumente und Kriterien im Vergleich lassen sich Tabelle B7 des elektronischen Zusatzmaterials entnehmen. Dort findet sich auch, wie viele Personen pro Klasse minimal und maximal eine Rolle einnehmen und in wie vielen Klassen die verschiedenen Rollen gar nicht auftreten. Es fällt auf, dass beim z -Kriterium unabhängig vom Instrument die Rollen in allen Klassen zu finden sind, während es beim min3-Kriterium auch Klassen gibt, in denen einzelne oder mehrere Rollen nicht identifizierbar sind. Generell ist aus den Prävalenzen in Tabelle 2.3 erkennbar, dass mit dem min3-Kriterium weniger Heranwachsenden eine Rolle zugeordnet werden kann.

Eine Analyse der offenen Antworten im RoleGrid kann eingeschränkt Aufschluss über diejenigen geben, die nicht oder nur selten (weniger als dreimal) nominiert wurden und somit nach dem min3-Kriterium keine Rolle haben. In einem freien Antwortfeld konnten die Schülerinnen und Schüler nach der Nominierungsprozedur für jede Form von Bullying angeben, was die übrigen Klassenmitglieder währenddessen machen. Diese Möglichkeit nahmen je nach Bullyingform 65 bis 104 der Befragten wahr. Insgesamt konnte ein Großteil der Antworten den Rollen zugeordnet werden. So entsprachen je nach Bullyingform

13 bis 28 Antworten der Verstärker-Rolle (z. B. lachen, zuschauen, dumme Kommentare, filmen), zwei bis acht Antworten der Verteidiger-Rolle (z. B. trösten, Hilfe holen, dazwischengehen, sagen, dass sie aufhören sollen) und sechs bis 15 der Außenstehenden-Rolle (z. B. kümmern sich nicht, ignorieren es, machen sich aus dem Staub, hören nicht zu, wegschauen). Ferner gab es 13- bis 20-mal die Antwort „nichts“, die ebenfalls der Außenstehenden-Rolle zuzuordnen wäre. Aus den übrigen Antworten wurde oftmals deutlich, dass nicht alle Klassenmitglieder involviert sind. Zum Teil wurde explizit geantwortet, dass die übrigen nichts damit zu tun haben (z. B. bekommen es nicht mit, wissen nichts, sind nicht dabei; bei sechs bis 14 Antworten), zum Teil wurden Aktivitäten genannt, die nichts mit den Schikanen zu tun haben (z. B. spielen, essen und trinken, Mittagspause, reden; bei 16 bis 33 Antworten).

Da für das RoleGrid beim min3-Kriterium eine Rollenzuordnung bereits auf Basis eines Indikators getroffen wird, ist von Interesse, wie viele Personen auch für mehrere Bullyingformen mindestens drei Nennungen erhalten haben. Eine entsprechende Übersicht gibt Tabelle 2.4. Es fällt auf, dass unabhängig von der Rolle die meisten das Kriterium tatsächlich nur für eine Bullyingform erfüllen. Nichtsdestoweniger werden gut 60 % der Opfer auf mehr als eine Weise schikaniert und etwa die Hälfte der Täter, Verteidiger und Außenstehenden sind für mehr als eine Form in dieser Rolle zu finden. Nur die Verstärker-Rolle beschränkt sich mit rund 80 % mehrheitlich auf eine einzelne Bullyingform.

Tabelle 2.4 Personenzahl, die das min3-Kriterium für eine bestimmte Anzahl an Formen von Bullying erfüllen

| Rolle laut RoleGrid | Anzahl der Formen mit mind. 3 Nennungen | | | | | <i>n</i> (mind. eine Form) |
|---------------------|-----------------------------------------|-----------|-----------|----------|----------|----------------------------|
| | 1 | 2 | 3 | 4 | 5 | |
| Opfer | 24 (37.5) | 18 (28.5) | 10 (15.6) | 7 (10.9) | 5 (7.8) | 64 |
| Bully | 85 (53.1) | 37 (23.1) | 21 (13.1) | 5 (3.1) | 12 (7.5) | 160 |
| Verstärker | 49 (79.1) | 7 (11.3) | 2 (3.2) | 2 (3.2) | 2 (3.2) | 62 |
| Verteidiger | 28 (46.7) | 20 (33.3) | 5 (8.3) | 4 (6.7) | 3 (5.0) | 60 |
| Außenstehende | 36 (51.4) | 13 (18.6) | 14 (20.0) | 6 (8.6) | 1 (1.4) | 70 |

Anmerkung. In der letzten Spalte ist unter *n* die Anzahl derjenigen angegeben, die bei der jeweiligen Rolle für mindestens eine Form drei Nennungen erhalten haben. Die prozentualen Häufigkeiten in Bezug auf diese zugrunde liegende Menge *n* sind in Klammern angeführt.

Ebenfalls interessant ist in diesem Zusammenhang die nähere Betrachtung, für welche der unterschiedlichen Bullyingformen das Kriterium von mindestens drei

Nennungen erfüllt wird und wie oft. Wie aus Tabelle 2.5 ersichtlich wird, ist das verbale Schikanieren die Bullyingform, an der – unabhängig von der Rolle – die meisten Personen beteiligt sind: Knapp 60 % der Opfer werden verbal schikaniert und auch über die Hälfte der Verstärker, Verteidiger und Außenstehenden haben diese Rolle im Kontext von verbalem Bullying. Allein von den Tätern sind mehr an isolierendem als an verbalem Bullying beteiligt.

Tabelle 2.5 Personenzahl, die das min3-Kriterium für die verschiedenen Formen von Bullying erfüllen

| Rolle laut RoleGrid | Bullyingform | | | | | <i>n</i> (mind. eine Form) |
|---------------------|--------------|-----------|------------|------------|-----------|----------------------------|
| | physisch | verbal | relational | isolierend | materiell | |
| Opfer | 28 (43.8) | 38 (59.4) | 33 (51.6) | 28 (43.8) | 16 (25.0) | 64 |
| Bully | 44 (27.5) | 69 (43.1) | 69 (43.1) | 91 (56.9) | 29 (18.1) | 160 |
| Verstärker | 27 (43.5) | 33 (53.2) | 14 (22.6) | 4 (6.5) | 9 (14.5) | 62 |
| Verteidiger | 30 (50.0) | 37 (61.7) | 18 (30.0) | 20 (33.3) | 9 (15.0) | 60 |
| Außenstehende | 38 (54.3) | 50 (71.4) | 28 (40.0) | 14 (20.0) | 3 (4.3) | 70 |

Anmerkung. In der letzten Spalte ist unter *n* die Anzahl derjenigen angegeben, die bei der jeweiligen Rolle für mindestens eine Form drei Nennungen erhalten haben. Die prozentualen Häufigkeiten in Bezug auf diese zugrunde liegende Menge *n* sind in Klammern angeführt.

2.4.6 Vergleich der Zuordnungskriterien

Nachdem im vorigen Kapitel die Prävalenzen der Rollen für zwei mögliche Zuordnungsstrategien berichtet wurden, sollen beide Kriterien an dieser Stelle miteinander verglichen werden. Für das RoleGrid lässt sich feststellen, dass ein Großteil derer, die das min3-Kriterium erfüllen, auch laut z-Kriterium die entsprechende Rolle zugewiesen bekommt (60–100 % je nach Rolle). Hundertprozentige Übereinstimmung zeigt sich bei der Verteidiger-Rolle. Die größten Diskrepanzen liegen bei der Täter-Rolle vor. Hier ist es offenbar so, dass einige Personen zwar drei Nennungen für mindestens ein Täter-Item erhalten, der Gesamtwert der Täter-Skala jedoch nicht über dem Klassenniveau liegt oder der Wert einer anderen Skala deutlich mehr über dem Klassenschnitt liegt. Ähnlich sind die Befunde für den PRQ: Hier erfüllen 77–95 % derer, die mindestens drei Nennungen für eine Rolle haben, auch das z-Kriterium.

Umgekehrt erweist sich das min3-Kriterium als strenger als das z-Kriterium: Von denjenigen, die das z-Kriterium erfüllen, haben im RoleGrid je nach Rolle nur 21–69 % auch mindestens drei Nominierungen für diese Rolle und ausschließlich für diese Rolle. Beim PRQ sind es 27–73 %. Die übrigen, die zwar das z-Kriterium für eine Rolle erfüllen, nicht aber das min 3-Kriterium, sind zum großen Teil rollenlos (haben für keine Rolle mindestens drei Nennungen erhalten) oder nicht eindeutig einer Rolle zuzuordnen (haben für mehrere Rollen mindestens drei Nennungen erhalten). Beim RoleGrid ist vor allem ersteres der Fall, beim PRQ eher letzteres.

2.4.7 Übereinstimmung von RoleGrid und PRQ

Zur Absicherung der Validität des neuen RoleGrids wurde sowohl die korrelative Übereinstimmung der Skalen als auch die Übereinstimmung hinsichtlich der Rollenzuordnung mit dem etablierten PRQ (Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a) überprüft. Die Korrelationen zwischen den Skalen beider Instrumente sind in Tabelle 2.6 dargestellt. Berichtet werden zum einen die allgemeinen Korrelationen wie auch die Within-Group-Korrelationen, das heißt die Korrelationen zwischen den klassenweise zentrierten Skalenwerten. Die Korrelationen der unzentrierten Skalen entspricht von der Logik der Rollenzuordnung anhand eines absoluten Cut-off-Kriteriums, während die Korrelation klassenweise zentrierter Skalen der Logik des z-Kriteriums nahekommt. Jeweils zusätzlich berichtet wird Kendalls Tau als Rangkorrelationskoeffizient, der sich für nicht-normalverteilte Daten und beim Vorliegen von Extremwerten gut bewährt hat (Howell, 2010; Leonhart, 2017). Das zusätzliche Trimmen der Daten hatte keinen weiteren Einfluss auf die Korrelationshöhe und wird daher nicht separat aufgeführt.

Tabelle 2.6 Korrelative Übereinstimmung zwischen RoleGrid und PRQ. (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a)

| RoleGrid | PRQ | | | |
|---------------|----------|--------------|-------------|---------------|
| | Opfer | Pro-Bullying | Verteidiger | Außenstehende |
| Opfer | .906*** | .000 | -.103* | .031 |
| | .585*** | .133*** | -.062 | .023 |
| | .914*** | -.035 | -.134** | .001 |
| | .549*** | .116*** | -.070* | .008 |
| Bully | .012 | .822*** | -.173*** | -.173*** |
| | .095** | .539*** | -.120*** | -.123*** |
| | -.034 | .838*** | -.219*** | -.282*** |
| | .111*** | .575*** | -.155*** | -.231*** |
| Verstärker | -.015 | .517*** | -.150*** | -.074 |
| | .114*** | .431*** | -.127*** | -.058 |
| | -.084* | .485*** | -.231*** | -.180*** |
| | .049 | .381*** | -.189*** | -.139*** |
| Verteidiger | -.084* | -.162*** | .733*** | .164*** |
| | -.025 | -.103** | .513*** | .208*** |
| | -.145*** | -.260*** | .753*** | .142*** |
| | -.000 | -.140*** | .461*** | .141*** |
| Außenstehende | .053 | -.102* | .002 | .526*** |
| | .057 | -.063* | .073* | .364*** |
| | .000 | -.215*** | .000 | .518*** |
| | .039 | -.204*** | .029 | .317*** |

Anmerkung. In der ersten Zeile steht jeweils Pearsons Korrelationskoeffizient r , in der zweiten Zeile Kendalls Rangkorrelationskoeffizient τ , in der dritten Zeile Pearsons r für die klassenweise zentrieren Mittelwertskalen und Items und in der vierten Zeile Kendalls τ für die klassenweise zentrierten Mittelwertskalen und Items; $N = 576$.

* $p < .05$, ** $p < .01$, *** $p < .001$

Unabhängig vom gewählten Zusammenhangsmaß zeigt sich die höchste Korrelation durchgängig positiv zwischen den beiden korrespondierenden Skalen von RoleGrid und PRQ (Salmivalli et al., 1996; deutsche Übersetzung von Schäfer & Korn, 2004a). Diese ist jeweils auf einem Niveau von $p < .001$ signifikant. Die übrigen Korrelationen sind überwiegend negativ, teilweise auch nicht signifikant. Eine Ausnahme ist die positive Rangkorrelation zwischen Täterbeziehungswise Pro-Bullying-Skala und Opfer-Skala, eine weitere Ausnahme die durchweg positive Korrelation zwischen Verteidiger-Skala des RoleGrid und Außenstehenden-Skala des PRQ. Grundsätzlich zeichnet sich über die verschiedenen Korrelationstypen ein recht einheitliches Bild von den Zusammenhängen.

Die Rangkorrelationen fallen mehrheitlich etwas geringer aus als Pearsons Korrelationen und die Within-Group-Korrelationen meist etwas höher. Am generellen Muster signifikanter Zusammenhänge ändert dies in der Regel nichts. Generell ähnelt das Korrelationsmuster zwischen den Rollen der verschiedenen Instrumente auch jenem Muster, das zwischen den Skalen eines Instrumentes erkennbar ist (siehe Tabelle B5 und B6 im Anhang, ESM).

Neben der korrelativen Übereinstimmung zwischen RoleGrid und PRQ wurde auch die Übereinstimmung hinsichtlich der Rollenzuordnung betrachtet. Legt man die von Noack und Petermann (1995) vorgeschlagenen Kenngrößen für die Beurteilung der Güte einer diagnostischen Entscheidung zugrunde, sind in diesem Zusammenhang zwei Fragestellungen relevant: Zum einen die Frage nach der Sensitivität des neuen Instruments (Wie viele derer, denen mittels PRQ eine Rolle zugewiesen wird, werden auch mittels RoleGrid für diese Rolle identifiziert?) und zum anderen nach dessen positivem Vorhersagewert (Wie viele derer, denen mittels RoleGrid eine Rolle zugewiesen wird, werden auch mittels PRQ für diese Rolle identifiziert?).

Prozentuale Angaben zur Sensitivität des RoleGrid unter Anwendung des min3-Kriteriums finden sich in Tabelle 2.7. Von denjenigen, denen mittels PRQ eine Rolle zugewiesen wird, erhalten auch 31–65 % durch das RoleGrid dieselbe Rolle. Den übrigen kann mittels RoleGrid größtenteils keine oder keine eindeutige Rolle zugeordnet werden. Nur bei acht Personen identifiziert das RoleGrid eine andere Rolle als der PRQ. Gut ist die Übereinstimmung insbesondere bei den Opfern (65 % der Opfer laut PRQ werden auch vom RoleGrid als solche erkannt) und bei den Pro-Bullying-Akteuren (54 % dieser werden auch vom RoleGrid als Täter oder Verstärker erkannt). Vielen Verteidigern und Außenstehenden laut PRQ wird hingegen vom RoleGrid keine Rolle zugewiesen.

Tabelle 2.7 Sensitivität des RoleGrid für die Rollen laut PRQ. (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a) bei Anwendung des min3-Kriteriums

| RoleGrid | PRQ | | | | | |
|---------------|----------------|----------------|----------------|----------------|----------------|-----------------|
| | Opfer | Pro-Bullying | Verteidiger | Außenstehende | uneindeutig | keine Rolle |
| Opfer | 65.0 % (13) | 0.0 % (0) | 1.6 % (1) | 0.0 % (0) | 12.4 % (21) | 1.4 % (2) |
| Bully | 5.0 % (1) | 43.2 % (54) | 3.1 % (2) | 1.9 % (1) | 11.8 % (20) | 9.7 % (14) |
| Verstärker | 0.0 % (0) | 11.2 % (14) | 0.0 % (0) | 1.9 % (1) | 2.4 % (4) | 2.8 % (4) |
| Verteidiger | 0.0 % (0) | 0.0 % (0) | 31.3 % (20) | 1.9 % (1) | 11.8 % (20) | 2.8 % (4) |
| Außenstehende | 0.0 % (0) | 1.6 % (2) | 0.0 % (0) | 30.8 % (16) | 7.1 % (12) | 5.5 % (8) |
| uneindeutig | 15.5 % (3) | 24.0 % (30) | 3.1 % (2) | 13.5 % (7) | 23.5 % (40) | 3.4 % (5) |
| keine Rolle | 15.5 % (3) | 20.0 % (25) | 60.9 % (39) | 50.0 % (26) | 31.2 % (53) | 74.5 % (108) |
| <i>n</i> | 20 | 125 | 64 | 52 | 170 | 145 |

Anmerkung. Prozentuale Angabe wie viele derer, denen mittels PRQ eine Rolle zugewiesen wird, auch mittels RoleGrid für diese Rolle identifiziert werden. Die Rollenzuweisung erfolgt bei beiden Instrumenten, wenn für mindestens ein Item der Rollenskala drei oder mehr Nennungen vorliegen. Die absoluten Häufigkeiten sind in Klammern angegeben; $N = 576$.

Die Angaben in Tabelle 2.8 ermöglichen es den positiven Vorhersagewert des RoleGrid für die Rollen des PRQ bei Einsatz des min3-Kriteriums einzuschätzen. Von denjenigen, denen mittels RoleGrid eine Rolle zugewiesen wird, erhalten auch 35–61 % durch den PRQ dieselbe Rolle. Den übrigen wird mittels PRQ größtenteils keine eindeutige oder gar keine Rolle zugeordnet. Am höchsten ist die Übereinstimmung bei den Tätern und Verstärkern des RoleGrid, die jeweils zu rund 60 % auch vom PRQ als Pro-Bullying-Akteure klassifiziert werden. Von den Opfern laut RoleGrid wird hingegen nur ein vergleichsweise geringer Anteil von 35 % auch mittels PRQ als Opfer identifiziert.

Tabelle 2.8 Positiver Vorhersagewert des RoleGrid für die Rollen laut PRQ (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a) bei Anwendung des min3-Kriteriums

| | PRQ | | | | | <i>n</i> |
|---------------|----------------|----------------|----------------|----------------|----------------|-----------------|
| | Opfer | Pro-Bullying | Verteidiger | Außenstehend | uneindeutig | |
| Opfer | 35.1 % (13) | 0.0 % (0) | 2.7 % (1) | 0.0 % (0) | 56.8 % (21) | 5.4 % (2) |
| Bully | 1.1 % (1) | 58.7 % (54) | 2.2 % (2) | 1.1 % (1) | 21.7 % (20) | 15.2 % (14) |
| Verstärker | 0.0 % (0) | 60.9 % (14) | 0.0 % (0) | 4.3 % (1) | 17.4 % (4) | 17.4 % (4) |
| Verteidiger | 0.0 % (0) | 0.0 % (0) | 44.4 % (20) | 2.2 % (1) | 44.4 % (20) | 8.9 % (4) |
| Außenstehende | 0.0 % (0) | 5.3 % (2) | 0.0 % (0) | 42.1 % (16) | 31.6 % (12) | 21.1 % (8) |
| uneindeutig | 3.4 % (3) | 34.5 % (30) | 2.3 % (2) | 8.0 % (7) | 46.0 % (40) | 5.7 % (5) |
| keine Rolle | 1.2 % (3) | 9.8 % (25) | 15.4 % (39) | 10.2 % (26) | 20.9 % (53) | 42.5 % (108) |

Anmerkung. Prozentuale Angabe wie viele derer, denen mittels RoleGrid eine Rolle zugewiesen wird, auch mittels PRQ für diese Rolle identifiziert werden. Die Rollenzuweisung erfolgt bei beiden Instrumenten, wenn für mindestens ein Item der Rollenskala drei oder mehr Nennungen vorliegen. Die absoluten Häufigkeiten sind in Klammern angegeben; $N = 576$.

Von der gesamten Stichprobe erhalten mit min3-Kriterium 46 % von beiden Instrumenten dieselbe Rollenzuweisung. Dass beide Instrumente unterschiedliche Rollen zuweisen, kommt jedoch nur in rund 1 % der Fälle vor. Mangelnde Übereinstimmung kommt vor allem zustande, wenn mit einem der Instrumente keine oder keine eindeutige Rolle identifiziert werden kann.

Unter Anwendung des z-Kriteriums liegt die Sensitivität des RoleGrid sogar bei 46–82 %. Der positive Vorhersagewert fällt mit je nach Rolle 46–79 % ebenfalls höher aus. Identische Rollen erhalten mit dem z-Kriterium von beiden Instrumenten 56 %. Allerdings steigt dabei auch die Anzahl der Personen, denen von den beiden Instrumenten verschiedene Rollen zugewiesen werden, drastisch auf 118 (20 %). Die vollständigen Daten zu Sensitivität und positivem Vorhersagewert bei Einsatz des z-Kriteriums finden sich im elektronischen Zusatzmaterial in den Tabellen B8 und B9.

2.4.8 Abgleich von Peer-Nominierungen mit Selbstbericht

Neben der Übereinstimmung des RoleGrid mit dem PRQ ist auch die Übereinstimmung zwischen Peer-Nominierungen und Selbstbericht relevant für die Beurteilung des neuen Instruments. Von den teilnehmenden 489 Schülerinnen und Schüler nannten sich im RoleGrid (jeweils mindestens für eine Bullyingform) 17 % als Opfer, 14 % als Täter, 8 % als Verstärker, 25 % als Verteidiger und 13 % als Außenstehende. Wird auf Basis der Selbstnominierungen eine Rollenzuordnung vorgenommen, wird deutlich, dass 45 % sich für gar keine Rolle und 16 % für mehrere Rollen selbst genannt haben. Diese Mehrfachrollen ausgenommen gibt es laut Selbstnominierungen im RoleGrid 9 % reine Opfer, 6 % reine Täter, 3 % reine Verstärker, 15 % reine Verteidiger und 7 % reine Außenstehende.

Je nach Rolle erhalten von denjenigen, die sich im RoleGrid selbst für eine Rolle nominieren, 11–42 % auch drei Fremdnominierungen. Hoch ist die Übereinstimmung für Täter und Opfer, gering für die Außenstehenden. Umgekehrt nennen sich von denjenigen, die im RoleGrid mindestens drei Fremdnominierungen erhielten, je nach Rolle auch 11–51 % selbst. Hoch ist die Übereinstimmung hier für Opfer und Verteidiger, gering wiederum für Außenstehende. Prüft man die Übereinstimmung auf Basis der Rollenzuweisung fallen die Übereinstimmungen geringer aus, da in beiden Fällen Mehrfachrollen möglich sind. Acht bis 30 % der selbstnominierten Rollen wird dieselbe Rolle mittels min3-Kriterium zugewiesen, 24–45 % mittels z-Kriterium. Interessant ist in diesem Zusammenhang, dass von jenen, die sich selbst keine Rolle oder die Rolle des Außenstehenden zuschrieben, rund 15 % als Täter gesehen wurden (min3-Kriterium). Mit

z-Kriterium wurden von denen, die sich für keine Rolle nannten, die meisten als Außenstehende (30 %) identifiziert, von denen, die sich selbst als Außenstehender sahen hingegen die meisten als Verstärker (27 %). Hervor sticht weiterhin, dass von den selbstnominierten Verstärkern 42 % von mindestens drei Mitschülern als Täter gesehen werden beziehungsweise 25 % das z-Kriterium für die Täter-Rolle aber auch 33 % das z-Kriterium für die Verteidiger-Rolle erfüllten. Betrachtet man, wie viele von denen, die laut min3-Kriterium eine Rolle zugewiesen bekommen, sich auch selbst für diese Rolle nennen, sind dies fünf bis 46 %. Besonders gering ist die Übereinstimmung bei Verstärkern, Außenstehenden und Tätern, hoch hingegen bei Opfern und Verteidigern. Das gleiche Muster zeigt sich mit z-Kriterium; hier nennen sich vier bis 40 % auch selbst. Die absoluten Häufigkeiten der selbstnominierten Rollen in Verbindungen mit den fremdnominierten Rollen sind in Tabelle 2.9 präsentiert.

Tabelle 2.9 Selbstenennungen und Fremdnominierungen für die Rollen im RoleGrid

| Rolle laut RoleGrid | Selbstnominierung | | | | | | | <i>n</i> |
|-----------------------|-------------------|-------|--------|----------|--------|---------|-------|----------|
| | Opfer | Bully | Verst. | Verteid. | Außen. | uneind. | keine | |
| min3-Kriterium | | | | | | | | |
| Opfer | 13 | 3 | 0 | 0 | 1 | 4 | 7 | 28 |
| Bully | 5 | 9 | 5 | 6 | 5 | 9 | 34 | 73 |
| Verstärker | 3 | 1 | 1 | 1 | 1 | 5 | 8 | 20 |
| Verteidiger | 2 | 1 | 2 | 15 | 3 | 8 | 11 | 42 |
| Außenstehende | 2 | 4 | 0 | 4 | 3 | 6 | 16 | 35 |
| uneindeutig | 4 | 7 | 0 | 14 | 7 | 15 | 35 | 82 |
| keine Rolle | 14 | 5 | 4 | 34 | 14 | 30 | 108 | 209 |
| z-Kriterium | | | | | | | | |
| Opfer | 17 | 3 | 0 | 2 | 3 | 7 | 11 | 43 |
| Bully | 3 | 10 | 3 | 5 | 4 | 10 | 27 | 62 |
| Verstärker | 5 | 5 | 3 | 8 | 9 | 21 | 33 | 84 |
| Verteidiger | 7 | 2 | 4 | 33 | 6 | 19 | 39 | 110 |
| Außenstehende | 6 | 7 | 0 | 13 | 8 | 12 | 65 | 111 |
| uneindeutig | 2 | 1 | 1 | 3 | 0 | 4 | 6 | 17 |
| keine Rolle | 3 | 2 | 1 | 10 | 4 | 4 | 38 | 62 |
| <i>n</i> | 43 | 30 | 12 | 74 | 34 | 77 | 219 | 489 |

Anmerkung. Absolute Häufigkeiten der auf Selbst- und Fremdnominierung beruhenden Rollen. Die Rollenzuweisung erfolgt bei der Selbstnominierung, wenn eine Person sich für mindestens einen Indikator der Rolle genannt hat. Die Rollenzuweisung erfolgt bei der Fremdnominierung mit min3-Kriterium, wenn eine Person für mindestens einen Indikator der Rolle mindestens dreimal genannt wurde, und mit z-Kriterium, wenn der Rollenscore einer Person über dem Klassenschnitt liegt ($z > 0$).

Die Anzahl der Bullyingformen, für die sich ein Schüler oder eine Schülerin selbst nominiert hat, korreliert bei der Opfer-, Täter- und Verteidiger-Rolle positiv mit der Anzahl an Bullyingformen, für die jemand mindestens dreimal genannt wurde (Kendalls $\tau = .302, .144$ und $.210$; $p < .001$). Keine signifikante Korrelation besteht hingegen bei der Rolle der Außenstehenden (Kendalls $\tau = -.026$; $p = .548$). Schwach ist die Korrelation für die Verstärker-Rolle (Kendalls $\tau = .092$; $p = .038$).

2.4.9 Zusammenhänge mit soziodemographischen und soziometrischen Merkmalen

Zur weiteren Beurteilung der Validität des neuen Instrumentes wurden die Zusammenhänge der Rollenskalen mit den soziodemographischen Merkmalen Geschlecht und Alter sowie mit den soziometrischen Merkmalen der Zuneigung (Liked-Most-Nominierungen) und Ablehnung (Liked-Least-Nominierungen) geprüft. Für einen ersten Überblick stellt Tabelle 2.10 zunächst die bivariaten Korrelationen vor. Hierzu werden wie in Abschnitt 2.4.4 und 2.4.7 neben Pearsons Korrelationskoeffizient auch Kendalls Rangkorrelation, die Within-Group-Korrelationen sowie die Korrelationen der getrimmten Skalen betrachtet. Kendalls Tau der getrimmten Skalen war abermals nahezu identisch mit Kendalls Tau der Rohwertskalen und wird daher nicht in der Tabelle aufgeführt.

Insgesamt zeigen sich die Korrelationen unabhängig von Zusammenhangsmaß und Datenaufbereitung relativ konsistent. Auch ist ein ähnliches Korrelationsmuster für die Skalen beider Instrumente mit soziodemographischen und soziometrischen Merkmalen erkennbar. Täter- und Verstärker- beziehungsweise Pro-Bullying-Skala sind mit dem männlichen Geschlecht assoziiert, Verteidiger- und Außenstehenden-Skala mit dem weiblichen. Auch für die Opfer-Skala besteht eine schwache Assoziation mit dem männlichen Geschlecht. Detailliertere Informationen zu Geschlechtsunterschieden (Mittelwertvergleiche, Rollenprävalenzen) finden sich in Tabelle B10 des elektronischen Zusatzmaterials.

Mit dem Alter zeigen sich keine deutlichen Zusammenhänge. Nur die nicht-zentrierte Täter- und Pro-Bullying-Skala ist leicht positiv mit dem Alter korreliert, ebenso die nicht-zentrierte Außenstehenden-Skala des PRQ. Innerhalb der Klassen zeigen sich bis auf eine Ausnahme keine Alterseffekte. Nur Klassenmitglieder, die älter als ihre Klassenkameraden sind, werden im PRQ seltener als Verteidiger genannt. Der Vergleich zwischen den Klassenstufen sechs und acht findet sich in Tabelle B11 des elektronischen Zusatzmaterials.

Die Opfer-Nominierungen hängen negativ mit Zuneigung und positiv mit Ablehnung zusammen. Die Täter-Nominierungen hängen ebenfalls positiv mit Ablehnung zusammen, gleichzeitig aber auch positiv mit Zuneigung – wenn auch schwächer ausgeprägt. Das gleiche Muster zeigt sich für die Pro-Bullying-Skala des PRQ. Für die Verstärker-Skala bestehen ebenfalls positive Korrelationen mit Zuneigung und Ablehnung, jedoch weniger konsistent über die verschiedenen Korrelationsmaße. Die Verteidiger-Nominierungen korrelieren hingegen durchweg positiv mit Zuneigung und negativ mit Ablehnung. Der Außenstehenden-Score zeigt schwache, inkonsistente Zusammenhänge mit den soziometrischen Items.

Tabelle 2.10 Korrelative Zusammenhänge der Rollen-Skalen mit soziodemographischen und soziometrischen Items

| Rolle | Geschlecht | Alter | Zuneigung | Ablehnung |
|-----------------|------------|-------|-----------|-----------|
| RoleGrid | | | | |
| Opfer | -.091* | .024 | -.219*** | .604*** |
| | -.076* | -.016 | -.111*** | .347*** |
| | -.090* | -.045 | -.307*** | .626*** |
| | -.056 | -.049 | -.161*** | .323*** |
| Bully | -.272*** | .102* | .168*** | .225*** |
| | -.227*** | .086* | .139*** | .218*** |
| | -.288*** | .030 | .096* | .216*** |
| | -.254*** | .002 | .074** | .212*** |
| Verstärker | -.332*** | .063 | .195*** | .076 |
| | -.344*** | .001 | .158*** | .155*** |
| | -.367*** | .008 | .080 | .043 |
| | -.358*** | .008 | .066* | .105*** |
| Verteidiger | .161*** | -.007 | .336*** | -.204*** |
| | .211*** | -.036 | .250*** | -.164*** |
| | .171*** | -.021 | .279*** | -.240*** |
| | .182*** | .003 | .189*** | -.166*** |
| Außenstehende | .125** | .019 | .021 | .005 |
| | .129*** | .021 | .070* | -.006 |
| | .118** | -.074 | -.102* | -.008 |
| | .127*** | -.059 | -.071* | -.019 |

(Fortsetzung)

Tabelle 2.10 (Fortsetzung)

| Rolle | Geschlecht | Alter | Zuneigung | Ablehnung |
|---------------|------------|--------|-----------|-----------|
| PRQ | | | | |
| Opfer | -.103* | .065 | -.212*** | .584*** |
| | -.079* | .109** | -.088** | .357*** |
| | -.105* | -.038 | -.348*** | .612*** |
| | -.053 | -.033 | -.198*** | .340*** |
| Pro-Bullying | -.271*** | .148** | .159*** | .273*** |
| | -.227*** | .094** | .144*** | .253*** |
| | -.285*** | .028 | .061 | .283*** |
| | -.221*** | .022 | .048 | .261*** |
| Verteidiger | .223*** | -.047 | .421*** | -.240*** |
| | .261*** | -.034 | .342*** | -.225*** |
| | .236*** | -.104* | .377*** | -.299*** |
| | .265*** | -.039 | .273*** | -.271*** |
| Außenstehende | .107* | .139** | .058 | -.031 |
| | .157*** | .094** | .039 | -.061 |
| | .128** | -.063 | -.023 | -.065 |
| | .150*** | -.046 | -.019 | -.065* |

Anmerkung. In der ersten Zeile steht jeweils Pearsons Korrelationskoeffizient r , in der zweiten Zeile Kendalls Rangkorrelationskoeffizient τ , in der dritten Zeile Pearsons r für die klassenweise zentrierten Variablen und in der vierten Zeile Kendalls τ für die klassenweise zentrierten Variablen; Geschlecht ist mit 0 für Jungen und 1 für Mädchen codiert; $N = 576$ bzw. für Korrelationen mit Alter aufgrund von Fehlwerten $N = 484$.

* $p < .05$, ** $p < .01$, *** $p < .001$

Um die bivariaten Zusammenhänge auch multivariat abzusichern, wurden für alle fünf Rollen des RoleGrid in R logistische Regressionsanalysen mit Random Intercepts gerechnet. Hierzu wurde die `glmer`-Funktion des `lme4`-Paketes mit der Spezifikation `family = binomial` genutzt. Dabei wurde jeweils die entsprechende Rolle laut `min3`-Kriterium der Gruppe der Rollenlosen (für kein Item mind. drei Nennungen) gegenübergestellt. Als Prädiktoren auf erster Ebene wurden Geschlecht, Zuneigung und Ablehnung aufgenommen. Zuneigung und Ablehnung wurden nicht – wie bei kontinuierlichen Variablen häufig üblich – zentriert, da die Liked-Most- und Liked-Least-Nennungen einen bedeutsamen Nullpunkt haben. Anstelle des Alters wurde auf die Klassenstufe als Prädiktor zurückgegriffen, um eine Reduktion der Stichprobengröße aufgrund von fehlenden Altersangaben ($n = 92$) zu verhindern. Dieses Vorgehen lässt sich gut rechtfertigen, da innerhalb der Klassen keine Zusammenhänge mit dem Alter bestehen (keine signifikanten bivariaten Within-Group-Korrelationen). Inhaltlich

entspricht dies zudem den in publizierten Studien häufig vorgenommenen Stufenvergleichen hinsichtlich der Prävalenz der Rollen. Auf zweiter Ebene wurde weiterhin die Teilnehmerzahl als Kontrollvariable berücksichtigt, diese zeigt jedoch bei keiner Rolle einen signifikanten Einfluss. Die Klassenstufe hat nur in Bezug auf die Opfer- und die Täter-Rolle einen signifikanten Effekt. Beide Rollen sind in den achten Klassen häufiger vertreten als in den sechsten (5.0 % Opfer in Stufe 6 vs. 8.0 % Opfer in Stufe 8, $B = 1.62$, $z = 3.11$, $p = .002$; 12.3 % Täter in Stufe 6 vs. 20.1 % Täter in Stufe 8, $B = 1.38$, $z = 2.62$, $p = .009$).

Auf individueller Ebene ist für die Opfer-Rolle nur die Ablehnung signifikant prädiktiv ($B = 0.63$, $z = 6.23$, $p < .001$). Die Täter-Rolle zeigt einen signifikant positiven Zusammenhang mit sowohl Ablehnung als auch Zuneigung – im Vergleich zu den Rollenlosen bekommen die Täter sowohl mehr Liked-Least- ($B = 0.43$, $z = 5.12$, $p < .001$) als auch mehr Liked-Most-Nennungen ($B = 0.33$, $z = 3.83$, $p < .001$). Zudem sind sie häufiger männlich ($B = -0.71$, $z = -2.62$, $p = .024$). Dies gilt auch für die Verstärker-Rolle ($B = -2.50$, $z = -3.23$, $p = .001$). Für die Verteidiger-Rolle ist in Abgrenzung zu den Rollenlosen nur die Zuneigung prädiktiv ($B = 0.39$, $z = 3.73$, $p < .001$). Keines der untersuchten Merkmale erweist sich als relevant für die Unterscheidung zwischen Außenstehenden und Rollenlosen. Alle Koeffizienten sowie die Odds-Ratios der Prädiktoren sind in Tabelle 2.11 aufgeführt.

Alle Modelle wurden hinsichtlich Multikolarität der Prädiktoren (mittels Variance Inflation Factor VIF und Toleranzstatistik) sowie Linearität zwischen Prädiktoren und Logit der Outcomevariablen geprüft (Field, Miles & Field, 2012). Es zeigten sich keine Voraussetzungsverletzungen. Zur Einschätzung der generellen Modellgüte wurden die Modelle für alle Rollen mit dem jeweiligen Nullmodell (Random-Intercept-Modell mit nur Teilnehmerzahl als Kontrollvariable) verglichen. Für Opfer-, Täter-, Verstärker- und Verteidiger-Rolle zeigte sich bei Hinzunahme der Prädiktoren Geschlecht, Zuneigung, Ablehnung und Stufe eine Abnahme des AIC (Akaike's information criterion) und eine signifikante Verbesserung des Modellfits (Opfer: $\chi^2_{(4)} = 91.65$, $p < .001$; Täter: $\chi^2_{(4)} = 54.54$, $p < .001$; Verstärker: $\chi^2_{(4)} = 21.33$, $p < .001$; Verteidiger: $\chi^2_{(4)} = 20.52$, $p < .001$; Außenstehende: $\chi^2_{(4)} = 5.37$, $p = .251$). Als standardisiertes Maß der multivariaten Assoziation zwischen festen Prädiktoren und beobachtetem Outcome wurde zudem mit der r2beta-Funktion des r2glmm-Paketes R_β^2 berechnet. Die höchsten Werte ergaben sich für die Modelle für Opfer-Rolle (.24) und Täter-Rolle (.22), gefolgt vom Modell für die Verteidiger-Rolle (.13) und zuletzt dem Modell für die Verstärker-Rolle (.09).

Tabelle 2.11 Logistische Regressionen mit Random Intercepts für die fünf Rollen des RoleGrid

| | Opfer | | | Bully | | | Verstärker | | | Verteidiger | | | Außenstehende | | |
|-----------------------|--------------------|---------------------|--|-------------------|---------------------|--|-------------------|---------------------|--|--------------------|---------------------|--|------------------|---------------------|--|
| | B (SE) | OR [95 % CI] | | B (SE) | OR [95 % CI] | | B (SE) | OR [95 % CI] | | B (SE) | OR [95 % CI] | | B (SE) | OR [95 % CI] | |
| Interzept | -5.37*** (1.20) | | | -2.15 (1.11) | | | -3.10 (2.36) | | | -4.87*** (1.30) | | | -6.79 (2.90) | | |
| Feste Effekte | | | | | | | | | | | | | | | |
| Ebene 1 | | | | | | | | | | | | | | | |
| Geschlecht | -0.04 (0.47) | 0.96 [0.38-2.45] | | -0.71* (0.32) | 0.49 [0.26-0.91] | | -2.50** (0.77) | 0.08 [0.02-0.37] | | 0.63 (0.41) | 1.88 [0.84-4.20] | | 0.92 (0.51) | 2.50 [0.92-6.80] | |
| LM | 0.19 (0.13) | 1.20 [0.93-1.56] | | 0.33*** (0.09) | 1.39 [1.18-1.65] | | 0.16 (0.16) | 1.17 [0.85-1.61] | | 0.39*** (0.11) | 1.48 [1.21-1.82] | | -0.14 (0.14) | 0.87 [0.66-1.14] | |
| LL | 0.63*** (0.10) | 1.87 [1.54-2.28] | | 0.43*** (0.08) | 1.54 [1.31-1.82] | | 0.27 (0.15) | 1.31 [0.97-1.75] | | -0.07 (0.58) | 0.93 [0.70-1.23] | | -0.00 (0.13) | 1.00 [0.78-1.28] | |
| Ebene 2 | | | | | | | | | | | | | | | |
| Stufe | 1.62** (0.52) | 5.05 [1.82-14.0] | | 1.38** (0.53) | 3.99 [1.42-11.2] | | 1.20 (1.15) | 3.33 [0.35-31.7] | | -0.19 (0.57) | 0.83 [0.27-2.56] | | 1.07 (1.22) | 2.92 [0.27-32.2] | |
| Teilnehmer- zahl | -0.00 (0.06) | 1.00 [0.89-1.11] | | -0.07 (0.05) | 0.93 [0.84-1.04] | | -0.05 (0.11) | 0.95 [0.76-1.18] | | 0.05 (0.06) | 1.05 [0.94-1.17] | | 0.15 (0.12) | 1.16 [0.92-1.48] | |
| Zufallseffekte | σ^2 (SD) | | | σ^2 (SD) | | | σ^2 (SD) | | | σ^2 (SD) | | | σ^2 (SD) | | |
| Interzept | 0.00 (0.00) | | | 0.93 (0.96) | | | 4.14 (2.04) | | | 0.73 (0.86) | | | 5.70 (2.39) | | |
| n | 291 (254 vs. 37) | | | 346 (254 vs. 92) | | | 277 (254 vs. 23) | | | 299 (254 vs. 45) | | | 292 (254 vs. 38) | | |

Anmerkung: Rollen wurden nach dem min3-Kriterium bestimmt, Referenzgruppe waren jeweils diejenigen, die dieses Kriterium für keine der Rollen erfüllen ($n = 254$). Geschlecht ist mit 0 für Jungen und 1 für Mädchen codiert; Stufe ist mit 0 für die 6. Klassen und 1 für die 8. Klassen codiert; LM = Anzahl der Liked-Most-Nennungen, LL = Anzahl der Liked-Least-Nennungen. Anzahl der Klassen: 23.
* $p < .05$, ** $p < .01$, *** $p < .001$

2.5 Diskussion

Die wesentlichen Zielsetzungen der Pilotierung des RoleGrid waren die generelle Erprobung des Erhebungsformates, die Entscheidung für ein sinnvolles Kriterium für die Rollenzuweisung sowie die Validierung des Instruments. Letzteres sollte anhand der Übereinstimmungsprüfung mit dem Participant Role Questionnaire (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a) sowie der Analyse von Zusammenhängen mit soziodemographischen und soziometrischen Merkmalen erfolgen.

2.5.1 Beantwortung der Fragestellungen

Insgesamt erwies sich das neu entwickelte RoleGrid als praktikables Instrument für die Erfassung der Rollen im Bullying-Geschehen. Im Folgenden soll nun detaillierter auf die einzelnen Teilfragen eingegangen werden.

(A) Feststellen der Praktikabilität des Erhebungsformates: Der Fragebogen wurde erfolgreich in 23 Klassen der Stufen sechs und acht von Werkreal-, Real- und Gemeinschaftsschulen durchgeführt, was die breite Einsetzbarkeit bestätigt. In Abhängigkeit von Klassenstufe und Schulform benötigten einige Klassen etwas mehr Zeit für die Bearbeitung oder eine intensivere Unterstützung. Insgesamt verliefen die Erhebungen reibungslos, der Begriff des Schikanierens und die Instruktionen schienen verstanden zu werden.

(B) Bestimmen von Itemkennwerten, faktorieller Struktur und Testgütekriterien: Die angenommene Fünf-Faktoren-Struktur des RoleGrid entsprechend der postulierten fünf Rollen ließ sich anhand von konfirmatorischen Faktorenanalysen bestätigen. Die approximativen Fitindizes sind zufriedenstellend und die Signifikanz des χ^2 -Tests ist vermutlich durch kleine Abweichungen in Kombination mit einer großen Stichprobe bedingt (Bühner, 2006; Tabachnick & Fidell, 2014). Insbesondere hervorzuheben ist, dass eine Differenzierung zwischen Tätern und Verstärkern mit dem RoleGrid besser zu gelingen scheint als mit dem PRQ (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a), obgleich die Faktoren dieser beiden Rollen hoch korrelieren. Die internen Konsistenzen (Cronbachs Alpha, McDonalds Omega) sind für alle Rollenskalen als gut zu bewerten. Dies passt zur Feststellung Olweus (2010), der hohe interne Konsistenz zwischen verschiedenen Bullyingformen als üblich bezeichnet. Die vorliegenden Befunde zeigen somit, dass die bisherigen Erkenntnisse zur internen Konsistenz von Selbstberichtsdaten zu Bullying und Viktimisierung auch für Peer-Nominierungen und auch für

die Bystander-Rollen gelten. Die Verteilungen von sowohl einzelnen Items als auch Skalenscores sind stark linkssteil und weisen eine große Anzahl an Nullen auf. Dies ist nicht ungewöhnlich für Peer-Nominierungen (Espelage, Holt & Henkel, 2003; Oldenburg et al., 2015; Velásquez et al., 2013) und insbesondere auch für extreme oder negativ konnotierte Merkmale (bspw. in klinischen Fragebögen, Bühner, 2006) und sollte daher nicht als Manko des Fragebogens gewertet werden. Dass dies im RoleGrid verglichen zum PRQ (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a) besonders extrem und auch bei den positiven Verhaltensweisen des Verteidigers auftritt, hat vermutlich mit der gezielten Gestaltung des neuen Fragebogens zu tun. Während der PRQ die unterschiedlichen Verhaltensweisen eher generell und unstrukturiert abfragt, steht beim RoleGrid am Anfang immer die Frage nach dem Opfer. So müssen sich die Befragten immer konkret Gedanken machen, wer genau dieses Opfer schikaniert oder wer genau diesem Opfer hilft. Die Wahrscheinlichkeit, dass Personen mit allgemein aggressivem oder allgemein prosozialem Verhalten genannt werden, sollte gerade durch diese Strukturierung reduziert werden. Die beobachteten Item- und Skalerverteilungen stellen somit zwar eine Herausforderung für die Auswertung dar, deuten gleichzeitig aber darauf hin, dass die befragten Schülerinnen und Schüler bei der Nominierung selektiver vorgegangen sind, was gerade Anliegen des neuen Instruments war. Letztlich sprechen faktorielle Struktur, interne Konsistenzen und auch die nicht-normale Verteilung somit für das Instrument.

(C) Festlegen eines geeigneten Kriteriums zur Identifikation der Rollen: Die Nachteile häufig eingesetzter Zuordnungskriterien für die Rollen wie beispielsweise einer Rollenzuweisung bei überdurchschnittlichen Skalenwerten (z-Kriterium) oder bei Überschreiten eines prozentualen Kriteriums wurden in Abschnitt 2.3.4 ausführlich beschrieben. Das z-Kriterium gewährleistet, wie sich auch in den vorliegenden Daten abzeichnet, dass dem Großteil der Heranwachsenden eine Rolle zugeordnet werden kann. Dies birgt jedoch die Gefahr falsch-positiver Rollenzuordnungen (Goossens et al., 2006). Ein prozentuales oder absolutes Kriterium ist demgegenüber strenger und bringt die Gefahr falsch-negativer Entscheidungen bei der Rollenzuordnung mit sich (Goossens et al., 2006). Insgesamt werden mehr Vorteile für ein absolutes Kriterium gesehen: Es ermöglicht eine leicht nachvollziehbare Identifikation der Rollen, die über verschiedene Klassen und Zeitpunkte hinweg Bestand hat und somit vergleichbar ist. Aufgrund der langfristigen Zielsetzung des Projektes, die Rollen hinsichtlich ihrer sozial-kognitiven und affektiven Reaktionen zu charakterisieren, erschien es zudem wichtiger falsch-positive Kategorisierungen zu vermeiden, um Gruppen mit möglichst eindeutigem Rollenverhalten zu erhalten. Das von Knauf, Eschenbeck und Käser (2017) eingesetzte absolute Kriterium von

mindestens sechs Nominierungen schien jedoch aus mehreren Gründen als zu hoch angesetzt: Zum einen ist das RoleGrid mit seiner sequentiellen Abfrage der verschiedenen Rollen darauf ausgelegt, leichtfertige Nominierungen zu verhindern. Die Bystander-Rollen werden für eine konkrete Opfer-Täter-Konstellation und eine konkrete Art von Bullying erfragt, sodass es insgesamt weniger Nominierungen gibt. Hinzu kommt, dass gerade für die vergleichsweise unauffälligen Verhaltensweisen der Bystander fraglich ist, wie viele Personen überhaupt davon mitbekommen. Die Grundlogik, eine Rolle erst dann zuzuordnen, wenn ein großer Teil der Klasse davon weiß und sich darin einig ist, wird daher als unpassend erachtet. Vielmehr sollte das absolute Kriterium darauf ausgelegt sein, Einzelaussagen durch weitere Aussagen abzusichern, um versehentliche oder mutwillige Falschnominierungen aussondern zu können. Die Wahl des Kriteriums von mindestens drei Nennungen stellt dabei einen Kompromiss aus den Validierungsstudien von Branson und Cornell (2009) sowie Phillips und Cornell (2012) dar. Drei Nominierungen stellen zudem für die meisten Items bereits einen Ausreißerwert dar, liegen also deutlich über der deskriptiven Norm. Somit erfüllt dieses Kriterium letztlich auch den Anspruch des z-Kriteriums und verschärft diesen sogar (vgl. Abschnitt 2.4.6). Dass damit die Zahl der Rollenlosen und der nicht eindeutig Klassifizierbaren hoch ausfällt, wirft wieder die Grundsatzfrage auf, ob tatsächlich die gesamte Klasse von den Vorfällen weiß und damit zu tun hat, ob tatsächlich in jeder Klasse alle Rollen vertreten sind (vgl. Pouwels et al., 2016) und inwiefern jedem eine eindeutige Rolle zukommt. Waasdorp und Bradshaw (2015) fanden mittels latenter Klassenanalyse heraus, dass 67 % der Zeugen von Bullying keinem der klassischen Reaktionsmuster von Bullying verstärkenden, verteidigenden oder passiven Verhaltensweisen zuzuordnen war, da sie eine geringe Wahrscheinlichkeit für alle abgefragten Verhaltensweisen oder ein inkonsistentes Muster berichteten. Werth et al. (2015) nennen zudem einige Forschende, die das Verhalten der Bystander von Bullying nicht als unveränderlich festgelegt, sondern als situationsabhängig variabel ansehen, wodurch auch Mehrfachrollen durchaus plausibel erscheinen. Das absolute Kriterium von mindestens drei Nominierungen wird somit – trotz der vergleichsweise geringen Anzahl an Heranwachsenden, denen damit eine klare Rolle zugeordnet werden kann – als Kriterium der Wahl erachtet.

(D) Abgleich mit PRQ und Selbstnominierungen: Zwischen den korrespondierenden Skalen von RoleGrid und PRQ (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a) zeigt sich eine klare korrelative Übereinstimmung. Auch die Korrelationen zwischen unterschiedlichen Rollenskalen zeigen ein plausibles Muster: Überwiegend sind die Zusammenhänge negativ, das heißt wer in einem

Instrument häufiger für eine Rolle genannt wurde, wird seltener für die anderen Rollen des anderen Instruments genannt. Ausnahme ist die positive Rangkorrelation zwischen Täter- und Opfer-Skala, welche sich jedoch auch innerhalb der Instrumente zeigt. Hinzu kommt eine positive Korrelation zwischen Verteidiger-Skala des RoleGrid und Außenstehenden-Skala des PRQ, welche sich möglicherweise durch die teils positiv konnotierte Formulierung der Außenstehenden-Items im PRQ (z. B. Item 21: Wer stellt sich auf keine Seite, wenn jemand schikaniert wird?) erklären lässt.

Die prozentuale Übereinstimmung beider Instrumente hinsichtlich der Rollenzuweisung fällt insgesamt moderat aus. Sensitivität und positiver Vorhersagewert des RoleGrid für die Rollen des PRQ sowie die Gesamtrate identisch zugewiesener Rollen sind bei Anwendung des z-Kriteriums etwas höher, was allerdings auch mit einer deutlich höheren Rate widersprüchlicher Rollenzuweisungen einhergeht. Widersprüchliche Rollenzuweisungen sind mit dem min3-Kriterium sehr selten. Mangelnde Übereinstimmung zwischen beiden Instrumenten rührt vielmehr daher, dass mit einem der beiden Instrumente keine oder keine klare Rolle identifiziert werden kann, während dies mit dem anderen Instrument möglich ist. Insgesamt sprechen die Befunde für die Validität des neuen Instrumentes. Eine hundertprozentige Übereinstimmung war weder zu erwarten, noch beabsichtigt, da das RoleGrid mit dem Ziel entwickelt wurde die Rollen – insbesondere die Bystander-Rollen – systematisch für die verschiedenen Bullyingformen zu erfassen. Weil dies im PRQ nur für die Opfer und teils auch für die Täter Berücksichtigung findet, ist es denkbar, dass die Nominierungen im RoleGrid akkurater für konkrete Bullying-Konstellationen sind, während die Nominierungen im PRQ möglicherweise auch allgemeine Verhaltenstendenzen (z. B. generell prosoziales oder passives Verhalten) widerspiegeln. Die Unterschiede zwischen den Zuordnungskriterien hinsichtlich der prozentualen Übereinstimmung beider Instrumente sprechen wiederum für die Nutzung des min3-Kriteriums, um Kategorien mit möglichst eindeutigem Rollenverhalten zu schaffen.

Im Abgleich der Peer-Nominierungen mit den Selbstnominierungen des RoleGrid zeigt sich mäßige Kongruenz. Dies mag zum Teil dadurch bedingt sein, dass die Befragten nicht daran gedacht haben sich selbst zu nominieren. Es steht jedoch auch im Einklang mit der generellen Feststellung Volks (2017), dass es zwar einen positiven Zusammenhang zwischen Selbst- und Fremdbericht gibt, die Übereinstimmung jedoch nicht unbedingt hoch ausfällt. Am höchsten ist die Übereinstimmung in dieser Studie für die Opfer-Rolle. Hingegen nominieren sich von den fremdnominierten Tätern und Verstärkern wenige selbst für diese Rolle, während von den selbstbenannten Verteidigern längst nicht alle auch von ihren Klassenkameraden als solche gesehen oder erkannt werden und von den selbstbenannten Verstärkern viele als

Täter nominiert werden. Dieses Muster zeugt von Selbstdarstellungstendenzen, wie sie auch von Salmivalli et al. (1996) berichtet werden, beziehungsweise allgemeiner formuliert von informantenspezifischen Anteilen in den Einschätzungen. Dies untermauert, dass beide Zugänge nicht redundant sind und dass Peer-Nominierungen einen eigenen wichtigen Beitrag in der Bullying-Forschung leisten (vgl. auch Badaly et al., 2015).

(E) Analyse der Zusammenhänge zwischen Rollenverhalten und soziodemographischen sowie soziometrischen Merkmalen: Die Validität des neuen Role-Grid sollte nicht nur anhand von PRQ (Salmivalli et al., 1996; deutsche Übersetzung nach Schäfer & Korn, 2004a) und Selbstauskünften, sondern zusätzlich durch externe Kriterien abgesichert werden. Geschlecht, Alter sowie soziometrischer Status haben in vergangenen Studien relativ klare Assoziationsmuster zu den Rollen im Bullying-Geschehen gezeigt, welche sich größtenteils auch in dieser Studie nachweisen lassen. Stabile bivariate Zusammenhänge in die erwartete Richtung zeigen sich zwischen Viktimisierung und geringer Akzeptanz sowie hoher Ablehnung, zwischen Täterschaft und dem männlichen Geschlecht und Ablehnung sowie – in geringerem Ausmaß – sozialer Akzeptanz, zwischen verstärkendem Verhalten und dem männlichen Geschlecht, zwischen verteidigendem Verhalten und dem weiblichen Geschlecht sowie hoher Zuneigung und geringer Ablehnung und schließlich zwischen passivem Verhalten und dem weiblichen Geschlecht. Schwach und nicht durchgängig signifikant sind die Assoziationen zwischen sowohl Verstärker- als auch Außenstehenden-Skala und dem soziometrischen Status. Tendenziell stehen die Befunde jedoch im Einklang mit der Erwartung, dass Verstärker den Tätern hinsichtlich des soziometrischen Status gleichen, jedoch näher am Durchschnitt liegen und dass Außenstehende weder hinsichtlich Zuneigung noch hinsichtlich Abneigung herausragende Werte erzielen. Kaum bivariate Zusammenhänge ließen sich wider Erwarten mit dem Alter feststellen, was jedoch an der relativ schmalen Alterspanne der Stichprobe liegen könnte. Das Muster der Korrelationen ist für RoleGrid und PRQ nahezu identisch, was die Validität des neuen Instruments untermauert.

Der multivariate Vergleich zwischen den fünf Rollen und jeweils der Gruppe der Rollenlosen zeigt, dass Opfer sich von den Rollenlosen insbesondere dadurch abzeichnen, dass sie stark abgelehnt werden und entgegen der Erwartung häufiger in den achten Klassen vorkommen. Die Täter sind im Vergleich zu den Rollenlosen eher männlich, sowohl beliebter als auch abgelehnter und kommen ebenfalls häufiger in den achten Klassen vor. Die Verstärker zeichnen sich gegenüber den Rollenlosen einzig durch ihr männliches Geschlecht aus. Die Verteidiger sind beliebter als die Rollenlosen und die Außenstehenden unterscheiden sich in keinem der untersuchten Merkmale von den Rollenlosen. Insgesamt entsprechen auch diese Befunde den

Erwartungen. Überraschend scheint zunächst, dass kein multivariater Zusammenhang zwischen Verteidiger-Rolle und Geschlecht zu bestehen scheint. Dies lässt sich jedoch damit erklären, dass die Vergleichsgruppe der Rollenlosen ebenfalls eher weiblich geprägt ist (50 % der Mädchen haben keine Rolle, während es bei den Jungen knapp 40 % sind). Deskriptiv gibt es mehr als doppelt so viele weibliche Verteidiger wie männliche Verteidiger. Weitgehend stehen die gefundenen bivariaten und multivariaten Assoziationen somit im Einklang mit den Hypothesen und sprechen für die Validität des RoleGrid.

Zusammenfassend lässt sich festhalten, dass das RoleGrid ein geeignetes Instrument ist, um die Rollen im Bullying-Geschehen unter Berücksichtigung der unterschiedlichen Schikaneformen zu erfassen. Die Erhebung war mit entsprechender Unterstützung ab Klassenstufe sechs gut machbar. Im Vergleich mit dem PRQ wird dem RoleGrid von einer geringeren Zahl an Schülern der Vorzug gegeben. Die Rollenzuweisung ist beim RoleGrid – insbesondere unter Verwendung eines absoluten Kriteriums von mindestens drei Nominierungen – selektiver als beim PRQ. Die angenommene Faktorenstruktur konnte bestätigt und gute interne Konsistenzen nachgewiesen werden. Auch die Differenzierung zwischen aktiven Tätern und Mit-tätern einerseits und deren Verstärkern andererseits ließ sich in der Faktorenstruktur des RoleGrid im Gegensatz zum PRQ eindeutig erkennen. Der Grad an Übereinstimmung mit dem PRQ spricht für die Validität des RoleGrid, zeigt aber auch, dass die beiden Instrumente nicht zu völlig identischen Ergebnissen kommen. Welches der beiden Instrumente das tatsächliche Rollenverhalten der Schülerinnen und Schüler besser abbildet, kann in dieser Studie nicht abschließend geklärt werden. Der strukturierte Aufbau des RoleGrid sollte dem PRQ jedoch insofern überlegen sein, dass er den kognitiven Prozess der Nominierung besser leitet, indem die Befragten bei jeder Rolle daran erinnert werden, dass diese auf konkrete aktuelle Fälle von Schikanen in der Klasse bezogen sind. Die Zusammenhänge mit externen Validierungskriterien wie dem Geschlecht sowie dem soziometrischen Status sprechen ebenfalls für Tauglichkeit des neuen Instrumentes.

2.5.2 Reflexion des methodischen Vorgehens

An dieser Stelle sollen die getroffenen methodischen Entscheidungen bezüglich der Gestaltung des Instruments sowie dessen statistischer Auswertung kritisch hinterfragt werden. Dies umfasst zum einen Fragen der Operationalisierung

des Rollenverhaltens (Nutzung von Peer-Report-Daten sowie Wahl des Zuordnungskriteriums) und zum anderen den Umgang mit den Besonderheiten der vorliegenden Daten (nicht-normalverteilte Zähl­daten mit genesteter Struktur).

Volk et al. (2017) bezeichnen die Messung von Bullying als Achillesferse der Bullyingforschung. Für die vorliegende Studie wurde daher penibel darauf geachtet, den Befragten den Begriff des Schikanierens verständlich zu machen und möglichst alle Erscheinungsformen von Schulbullying abzudecken, um eine hohe Validität zu gewährleisten. Obgleich es wie Volk et al. (2017) feststellen eine langandauernde Debatte über die Nützlichkeit von Selbstbericht- versus Peer-Report-Daten gibt, kommt diese Forschungsgruppe zu dem Schluss, dass mit den unterschiedlichen Zugängen auch Unterschiedliches gemessen wird und dass die Wahl des Informanten insbesondere vom logischen Modell der Studie abhängen sollte. Badaly et al. (2015) argumentieren, dass Selbst- und Fremdeinschätzungen sowohl geteilte als auch spezifische Informationen enthalten und daher beide Maße nützlich und berechtigt sind. Sowohl Selbstbericht als auch Peer-Report haben – wie auch die Befragung von Lehrern oder Eltern – Vor- und Nachteile. Über Peer-Auskünfte mögen zwar nicht alle Fälle von Bullying erfassbar sein, da es oftmals verdeckt abläuft und die Peer-Gruppe nicht über alles Bescheid weiß (Olweus, 2010; Volk et al., 2017), andererseits können mittels Peer-Auskünften Verzerrungen durch eine positive Selbstdarstellung umgangen und Fehleinschätzungen einzelner Personen ausgeglichen werden (Badaly et al., 2013). Für die Erfassung der Rollen im Bullying-Geschehen ist die Nutzung von Peer-Report-Daten besonders angemessen, da nur dies dem Konzept einer sozialen Rolle gerecht werden kann. Eine Rolle ergibt sich aus den Erwartungen, die an ein Individuum gerichtet sind und auf die auch das Interaktionsverhalten in der Gruppe abgestimmt ist (Six, 2020). Dementsprechend sollte die Identifikation der Rollen auf der Wahrnehmung der Gruppenmitglieder basieren und nicht ausschließlich auf einer Selbsteinschätzung. Auch Olweus (2010) sieht den Zweck von Peer-Nominierungen darin, typische Verhaltensmuster abzubilden, was dem Rollengedanken entspricht. Hinzu kommt perspektivisch gedacht, dass mit dem Einsatz von Fremdnominierungen für die Rollenerfassung einer Überschätzung von Zusammenhängen durch geteilte Methodenvarianz entgegengewirkt werden kann (Podsakoff, MacKenzie & Podsakoff, 2012). Für die Hauptstudie ist gerade von Interesse, inwiefern selbstberichtete sozial-kognitive und affektive Reaktionen auf Bullying mit beobachtetem Rollenverhalten einhergehen. Beachtet werden muss dabei, dass die Anzahl an Nominierungen nicht direkt Auskunft über Häufigkeit oder Intensität des Verhaltens gibt, sondern in erster Linie über dessen Sichtbarkeit (vgl. Olweus, 2010).

Diese Grundüberlegung ist auch relevant für die Auswahl des Zuordnungskriteriums. Es wurde von der Annahme abgesehen, dass alle Rollen für einen Großteil der Klasse sichtbar sind. Vielmehr wird davon ausgegangen, dass die rollentypischen Verhaltensweisen je nach Art des Bullying und auch je nach Rolle mehr oder weniger sichtbar sind. Nichtsdestoweniger wurde daran festgehalten, dass ein Verhaltensmuster zumindest für einige Klassenmitglieder sichtbar sein muss, damit es als Rolle bezeichnet werden kann. Um den Vorteil der Peer-Report-Methode ausnutzen zu können, sich nicht auf die Aussagen von Einzelpersonen verlassen zu müssen (Badaly et al., 2013), sollten sich mindestens drei Personen in Ihrer Einschätzung einig sein, damit eine Rolle zugeordnet wurde. Im Sinne der Einigkeit wurde der Trennwert von drei Nennungen auf die einzelnen Items angewendet und nicht auf den Skalenmittelwert, welcher auch erhöht sein kann, wenn die Nennungen für alle Items der Skala von einer einzigen Person stammen. Auch bei Oldenburg et al. (2015) werden Heranwachsende als Opfer betrachtet, wenn sie für mindestens eine Form der Viktimisierung genannt wurden. Ein absoluter Cut-off-Wert wurde gewählt, damit – im Gegensatz zur klassenweisen Standardisierungsmethode (z-Kriterium) – ein Vergleich zwischen verschiedenen Klassen und die Abbildung individueller Veränderungen möglich ist und damit – im Gegensatz zu prozentualen Cut-off-Werten – jeder Stimme das gleiche Gewicht zukommt. Die Logik der beiden alternativen Zuordnungsmethoden, dass ein Rollenverhalten über dem Klassenschnitt liegen muss oder von einem bestimmten Anteil der Klasse beobachtbar sein muss, wird nicht geteilt. Weder die deskriptive Klassennorm noch die Klassengröße sollten ausschlaggebend für die Identifikation der Rollen sein.

Ergebnis der Peer-Nominierungen sind nicht-normalverteilte Zähldaten mit einer großen Anzahl an Nullen einerseits und einigen Extremwerten andererseits. Diesen Besonderheiten wird in der Bullyingforschung wenig Beachtung geschenkt und die Daten meist mit den üblichen Verfahren für normalverteilte kontinuierliche Daten analysiert (Branson & Cornell, 2009; Phillips & Cornell, 2012; Schäfer & Korn, 2004a). Tabachnick und Fidell (2014) sowie auch Howell (2010) argumentieren, dass es auf das angenommene Skalenniveau der zugrundeliegenden latenten Variable ankommt. Tabachnick und Fidell (2014) weisen zudem darauf hin, dass auch die Anzahl richtig gelöster Aufgaben in einem Test oder die Antworten auf einer Ratingskala nicht genuin kontinuierlich sind, jedoch häufig als solche behandelt werden können, wenn viele Abstufungen vorliegen. Sie weisen allerdings auch darauf hin, dass die Verteilungsform das eigentlich bedeutsame Merkmal ist. Normalverteilte Fehler sind bei Zähldaten

unwahrscheinlich (O'Hara & Kotze, 2010) und tatsächlich sind auch die vorliegenden Daten nicht normalverteilt. Von Transformationen zur Normalisierung von Zähldaten wird abgeraten (O'Hara & Kotze, 2010).

Für regressionsanalytische Zwecke kann die Verteilungsform (z. B. Poisson, quasi-Poisson oder negativ-binomial) im Rahmen generalisierter linearer Modelle berücksichtigt werden (O'Hara & Kotze, 2010), für exploratorische sowie konfirmatorische Faktorenanalysen in SPSS oder R liegen hingegen keine entsprechenden Lösungen vor. Der für nicht-normalverteilte Daten empfohlene ADF-Schätzer (asymptotical distribution free), auch full WLS (weighted least squares) genannt, benötigt selbst für wenig komplexe Modelle enorm große Stichprobenumfänge (Bühner, 2006; Rosseel, 2012). Eine umfassende Simulationsstudie von Flora und Curran (2004) zeigt in Einklang mit vorausgehenden Studien, dass mit full-WLS-Schätzung bei großen Modellen mit 10 oder mehr Indikatoren und einer Stichprobengröße von unter $N = 1000$ die χ^2 -Statistik und die Parameterschätzungen überhöht und die Schätzungen der Standardfehler negativ verzerrt sind. Stärkere Abweichungen der beobachteten Variablen von der Normalverteilung gehen mit stärkeren Verzerrungen einher (Flora & Curran, 2004). Robuste WLS-Schätzer (weighted least squares) hingegen resultieren über alle Stichprobengrößen und für unterschiedlich komplexe Modelle in akkuraten Teststatistiken, Parameterschätzungen und Standardfehlern (Flora & Curran, 2004). Allerdings beruhen robuste WLS-Schätzer auf polychorischen Korrelationen (Schmitt, 2011), welche von normalverteilten latenten Variablen ausgehen (Kline, 2011; Li, 2016), was für das Rollenverhalten im Bullying-Geschehen wenig plausibel erscheint. Flora und Curran (2004) finden, dass polychorische Korrelationen robust gegenüber leichten bis moderaten Abweichungen der latenten Variablen von der Normalverteilung sind, nicht aber bei starken Abweichungen. Massive Schiefe und Kurtosis führen zu substanziell verfälschten Schätzungen der Korrelation und machen sie damit unbrauchbar für konfirmatorische Faktorenanalysen (Flora & Curran, 2004). Für den Fall, dass die zugrundeliegende Verteilung nicht-normal ist, die meisten Angaben in eine Kategorie fallen und es vor allem um die Evaluation des Modellfits geht, empfiehlt Schmitt (2011) robuste ML-Schätzer (maximum likelihood) in Betracht zu ziehen. Der robuste ML-Schätzer im lavaan-Paket gibt korrigierte Standardfehler sowie die Satorra-Bentler-skalierte Teststatistik aus, welche die χ^2 -Statistik entsprechend dem Ausmaß der Kurtosis korrigiert (Rosseel, 2012). Obgleich dieses Vorgehen den Besonderheiten der Daten nicht vollständig Rechnung trägt, scheint es die passendste Lösung für die hiesige Zielsetzung. Die Stichprobengröße für die konfirmatorische Faktorenanalyse lag mit 576 Schülerinnen und Schülern etwas unter dem empfohlenen Minimum von 10-mal mehr Probanden

als zu schätzende Parameter (Kline, 2011, S. 12). Für das Modell mit fünf korrelierten Faktoren und 25 Variablen müssen 60 Parameter geschätzt werden, wozu mindestens 600 Probanden hätten herangezogen werden sollten.

Da die Verteilungseigenschaften der beobachteten Werte auch für die Schätzung der internen Konsistenzen eine Rolle spielen (Liu & Zumbo, 2007), wurden die Konfidenzintervalle mittels BC_a -Bootstrapping (bias-corrected and accelerated) berechnet, welches eine Korrektur für Schiefe und Kurtosis vornimmt (Dunn et al., 2014). Um den Einfluss der Extremwerte auf sowohl Faktorenstruktur als auch interne Konsistenzen auszumachen, wurden die Analysen mit auf maximal drei Nennungen getrimmten Itemscores wiederholt, was nicht zu wesentlich abweichenden Ergebnissen führte.

Auch für die Überprüfung der Zusammenhänge zwischen den Rollenskalen der beiden Instrumente und den soziodemographischen und soziometrischen Merkmalen wurde neben Pearsons Korrelationskoeffizient auch Kendalls Rangkorrelation Tau berichtet. Der Vergleich zeigt, dass besonders die Korrelationen von Opfer- und Verteidigerskala mit Zuneigung und Ablehnung von Extremwerten beeinflusst wird. Dies bedeutet aber auch, dass die Varianz auf diesen Merkmalen durch das Trimmen besonders stark beschnitten wird. Da es sich bei den Extremwerten nicht um fälschlicherweise hohe Werte handelt, gibt das Trimmen nicht notwendigerweise die tatsächlichen Zusammenhänge besser wieder. Kendalls Tau wird für ordinalskalierte Daten mit Verbundrängen empfohlen (Bortz, Barskova, Leitner, Lienert & Oesterreich, 2008; Leonhart, 2017) und ist besser für stark nicht-normalverteilte Daten geeignet (Howell, 2010) als Pearsons Korrelation. Nichtsdestoweniger wird aus Gründen der Vergleichbarkeit mit anderen Publikationen auch Pearsons Produkt-Moment-Korrelation berichtet, welcher laut Howell (2010) der Korrelationskoeffizient der Wahl bleibt. Zudem liefert Pearsons Koeffizient ein deskriptives Maß des linearen Zusammenhangs, welches unabhängig von Verteilungsannahmen ist, und auch die Signifikanztestung hat sich als robust gegenüber Verletzungen der Verteilungsannahme sowie des Intervallskalenniveaus erwiesen (Bortz & Schuster, 2010).

Neben den bivariaten Zusammenhängen zwischen dem von den Klassenkameraden beobachteten Rollenverhalten und Geschlecht, Alter und soziometrischen Merkmalen wurden logistische Regressionen gewählt, um diese Zusammenhänge auch multivariat abzusichern. Diese wurden generalisierten linearen Modellen mit negativ binomialer Verteilungsannahme für Zähldaten aus drei Gründen vorgezogen: Zum einen entspricht die Analyse der Rollen als binärem Outcome dem grundsätzlichen Rollengedanken und wird somit dem Ziel, die Rollen zu charakterisieren besser gerecht als die Arbeit mit den Rollenskalen, welche das Ausmaß der Sichtbarkeit des Rollenverhaltens wiedergeben. Die Annahmen

beziehen sich in der Regel darauf, dass Personen, die ein Rollenverhalten zeigen, gewisse Merkmale aufweisen und weniger darauf, dass diese Merkmale umso stärker ausgeprägt sind, je mehr Personen das Rollenverhalten beobachten konnten. Des Weiteren kann somit wiederum die Validität des Zuordnungskriteriums von mindestens drei Nominierungen geprüft werden, da die Dichotomisierung des Outcomes anhand dieses Kriteriums erfolgte. Darüber hinaus hat dieses Vorgehen den Vorteil, dass nur die eindeutigen Rollen – jeweils im Vergleich mit den Rollenlosen – betrachtet werden, während bei der Analyse der Rollenskalen nicht ausgeschlossen ist, dass einige Personen mit hohem Score für eine Rolle nicht auch hohe Werte für andere Rollen haben. Ob jene Personen, denen keine Rolle zugeordnet werden konnte, die am besten geeignete Vergleichsgruppe darstellen, mag hinterfragt werden, da diese letztlich recht heterogen zusammengesetzt sind. Diese Gruppe umfasst (1) Personen, die sich in Bezug auf das Bullying-Geschehen so subtil verhalten, dass sie gar nicht genannt werden, (2) Personen, bei denen nur selten typisches Rollenverhalten erkennbar ist und somit nur ein bis zwei Nennungen für eine Rolle erhalten, (3) Personen, deren Verhalten wechselhaft und nicht klar einzuordnen ist und die daher möglicherweise für verschiedene Rollen ein bis zwei Nennungen haben sowie (4) prinzipiell auch Personen, die keine Rolle einnehmen, weil in dieser Klasse kein Bullying stattfindet. Für folgende Analysen wäre es möglich, die Gruppe der Rollenlosen zu homogenisieren, indem beispielsweise nur jene mit genau null Nennungen oder nur jene aus Klassen, in denen gar keine Rollen vergeben wurden, als Vergleichsgruppe heranzuziehen. Auch die Nutzung der Außenstehenden als Referenzgruppe ist denkbar, die sich in dieser Studie nicht von den Rollenlosen unterscheiden, inhaltlich jedoch klarer zu verorten sind. Vorteil der Arbeit mit den Rollengruppen ist, dass grundsätzlich jedwede Rollenpaare gegeneinander kontrastiert werden können, deren Vergleich inhaltlich für eine bestimmte Fragestellung interessiert.

Eine weitere Besonderheit der Daten, die sich nicht per se durch die Methode der Peer-Nominierung als vielmehr durch die Erhebung in Klassenverbänden ergab, ist die genestete Struktur der Daten, welche sich in erhöhten Designeffekten zeigte. Um die Niveauunterschiede zwischen den Klassen zu eliminieren, wurden die Itemscores und die Rollenscores für die Berechnung von Faktorenanalysen, Reliabilitäten und Korrelationen klassenweise zentriert. Der Vergleich mit Berechnungen für die nicht-zentrierten Werte zeigte nur geringe Unterschiede. Diskutabel bleibt nichtsdestoweniger, ob eine klassenweise Zentrierung der Daten für die Analyse der Faktorenstruktur überhaupt angemessen ist. Zum einen wird eine Transformation vorgenommen, welche den natürlichen Nullpunkt der Skala

und zwischen den Klassen und Items auch die natürliche Einheit unterläuft². Zum anderen werden Niveauunterschiede zwischen den Klassen und Items nivelliert, die nicht ausschließlich artifiziell sind, sondern vermutlich zu großen Teilen auch wahre Unterschiede widerspiegeln (vgl. Reflexion der Rollenzuordnung anhand von klassenweise z-standardisierten Werten in Abschnitt 2.3.4). Faktorenanalysen mit klassenweise zentrierten Werten resultieren schlichtweg in einer Aussage über die Zusammenhangsstruktur verschiedener Bullying-bezogener Verhaltensweisen *innerhalb* der Klassen, während Analysen mit unzentrierten Daten die Zusammenhänge in der Gesamtstichprobe abbilden. Eine Analyse mit Rohdaten entspricht letztlich eher dem bevorzugten absoluten Kriterium zur Rollenzuordnung. Für die Regressionsanalysen konnte die elegantere Lösung mit Random-Intercept-Modellen zur Berücksichtigung der genesteten Datenstruktur gewählt werden.

Zusammenfassend stellen Peer-Nominierungs-Daten, die im Klassenverbund erhoben werden müssen, eine Reihe an Herausforderungen an deren Auswertung. Die Nutzung von generalisierten linearen Modellen anstelle von gaußschen Regressionsmodellen sowie die Berücksichtigung der Mehrebenenstruktur hält zunehmend Einzug in die Bullyingforschung, ein Goldstandard für die Auswertung hat sich bislang jedoch nicht herauskristallisiert. Zudem mangelt es an frei zugänglicher Software, mit der verschiedene Datentypen und Verteilungsformen sowie die Mehrebenenstruktur nutzerfreundlich berücksichtigt werden können.

2.5.3 Ausblick

Aus der vorliegenden Pilotstudie lassen sich Implikationen für die Hauptstudie, aber auch ein genereller Ausblick für die künftige Forschung herleiten. Zunächst lassen sich praktische Schlüsse für die Gestaltung des Fragebogens ziehen, die zu leichten Modifikationen des Fragebogens für die Hauptidehebung führen. Die erste Spalte für die Opfer-Nominierungen sollte aufgrund wiederholter Nachfragen der Studienteilnehmer eindeutiger beschriftet werden. Für die Selbstnominierungen sollte es eine Ankreuzmöglichkeit geben, damit die Befragten nicht vergessen,

² Beispiel: Klassenmittelwert für ein Item ist in einer Klasse 0.91 und in einer anderen 2.18. Mit vier Nominierungen erhält man in der ersten Klasse einen klassenzentrierten Wert von 3.09 und mit drei Nennungen einen Wert von 2.09. Die Differenz von einer Nennung bleibt innerhalb der Klassen erhalten. In der zweiten Klasse hingegen erhält man mit drei Nennungen einen klassenzentrierten Wert von 0.82. Die Differenz von einer Nennung zwischen den Klassen wird durch die Differenz der klassenzentrierten Werte ($3.09 - 0.82 = 2.27$) nicht abgebildet.

auch zu sich selbst ein Urteil abzugeben. Die Frage danach, was die übrigen Klassenmitglieder während der Schikanen tun, war zwar informativ für die Pilotstudie, schien die Befragten jedoch dazu zu verleiten keine differenzierten Nominierungen für alle Einzelpersonen vorzunehmen. Möglicherweise wurden selektiv nur diejenigen benannt, deren Rollenverhalten besonders salient war, und die übrigen pauschal mit der letzten Frage abgedeckt. Die gegebenen Antworten ließen sich fast durchgängig den erfragten Bystander-Rollen (Verstärker, Verteidiger, Außenstehender) zuordnen. War dies nicht der Fall, ließen die Antworten darauf schließen, dass nicht alle Personen mit den Schikanen zu tun hatten. Es gab keinerlei Hinweise dafür, dass bedeutsame Verhaltensmuster im Bullying-Geschehen übersehen wurden und im RoleGrid fehlen. Die offene Frage sollte dementsprechend für Folgeuntersuchungen entfallen, um die Befragten dazu anzuhalten, sich über alle Klassenmitglieder separat Gedanken zu machen und diese nicht als „die übrigen“ zusammenzufassen. Um die Einordnung zu erleichtern und die Rollen noch klarer zu machen, sollten diese durch vorangehende Beschreibungen kurz skizziert werden (vgl. Oldenburg et al., 2015), auf welche die Schülerinnen und Schüler bei der Befragung immer wieder zurückgreifen können.

Neben diesen pragmatischen Folgerungen aus der Pilotstudie, wäre eine Validierung des RoleGrid mittels Interviews und Beobachtungen wünschenswert. Die Opfer-Rolle könnte ähnlich wie bei Phillips und Cornell (2012) durch Gespräche mit Schulpsychologen oder Schulsozialarbeitern bestätigt werden. Weiterhin haben Wójcik und Mondry (2020) über Beobachtung im Unterricht, vor allem aber in den Pausen sowie vor und nach dem Unterricht, verschiedene Rollen im Bullying-Geschehen identifizieren können. Ein solches Beobachtungsschema wäre höchst aussagekräftig für den Abgleich mit den via RoleGrid identifizierten Rollen.

Insgesamt sind die Befunde ermutigend für das Anliegen, das RoleGrid auch um Items zu Cyberbullying zu ergänzen. Verglichen mit Selbstberichtfragebögen wird die Methode der Peer-Nominierung in der Bullyingforschung seltener eingesetzt (Cook, Williams, Guerra & Kim, 2010), findet jedoch vor allem im Zusammenhang mit den Participant Roles ein sinnvolles Anwendungsfeld. Publierte Versuche, auch das Bystander-Verhalten bei Cyberbullying über Peer-Nominierungen zu erfassen, scheint es bislang jedoch nicht zu geben. Die Forschungsgruppe um Badaly (Badaly et al., 2013; Badaly et al., 2015) hat Peer-Nominierungen bereits erfolgreich für die Erfassung von elektronischer Aggression und elektronischer Viktimisierung eingesetzt und regen an Peer-Nominierungs-Instrumente in Forschungsbestrebungen zur digitalen Welt einzubeziehen. Anliegen der Hauptstudie ist daher eine Erweiterung des RoleGrid um Items zu Cyberbullying, um neben Tätern und Opfern auch Verstärker,

Verteidiger und Außenstehende bei Cyberbullying über Peer-Auskünfte zu identifizieren. Es sollen die Zusammenhänge aufgedeckt werden zwischen diesen beobachtbaren Verhaltensmustern und den selbstberichteten mentalen Reaktionen auf Schul- und Cyberbullying. Ein solches Instrument ermöglicht weiterhin eine Untersuchung, inwiefern die Verhaltensmuster kontextübergreifend feststellbar sind, wie hoch die Überlappungen zwischen verschiedenen Bullyingformen sind und ob es möglicherweise Bullyingform-spezifische Assoziationsmuster gibt.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

