

# **The authenticity of simulated patients in psychotherapy training and research**

DISSERTATION  
zur Erlangung des akademischen Grades

Doctor philosophiae  
(Dr. phil.)

im Fach Psychologie

eingereicht bei der Humanwissenschaftlichen Fakultät der  
Universität Potsdam

von

**Destina Sevde Ay-Bryson**

2021

Erstgutachter: Prof. Dr. Florian Weck  
Zweitgutachterin: Prof. Dr. Eva-Lotta Brakemeier  
Datum der Disputation: 07.01.2022

Published online on the  
Publication Server of the University of Potsdam:  
<https://doi.org/10.25932/publishup-53373>  
<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-533735>

## Acknowledgements

I am infinitely grateful to both my supervisors, Florian Weck and Franziska Kühne, who have been greatly strengthening, motivating and inspiring throughout my doctoral studies. I benefited immensely from their guidance both scientifically as a graduate student and clinically as a psychotherapy trainee. Many thanks for your time, dedication and your investment in me.

During my doctoral studies, I was a scholarship holder of the Hans-Böckler-Stiftung. I am greatly thankful for their financial and ideal support for funding this project with a full scholarship. Further, the University of Potsdam (Koordinationsbüro für Chancengleichheit) granted me a scholarship during the planning phase of my dissertation as part of the “Brückenprogramm Chancengleichheit”, for which I am also very grateful.

I thank all of my colleagues at the division of Klinische Psychologie und Psychotherapie der Universität Potsdam, especially Ulrike Maaß, Peter E. Heinze, Ewa Gartenschläger, Arne Klopffleisch and Maik Schellhase for your highly supporting and positive natures. I greatly appreciated being a member of this fantastic team.

I thank my family. I am forever thankful to Necla Ay for being the greatest of role models of a strong woman and dearly-loving mother – you are my personal heroine and I wouldn't be where I am were it not for you. I am thankful to Adnan Ay for being my greatest fan and the most generous father. My grandmother, Aysel Duymaz, has shown me that age doesn't stop you from learning and personally growing – thank you so much for our special and caring friendship and for teaching me how to be independent and to follow my own nose. I thank my sister, Eda Ay, for always believing in me and for spreading so much empathy into the world.

I thank all of my friends who have been there for and with me, from drunken laughter to kitchen philosophy, from Game of Thrones nights to culinary happenings, from hikes to book clubs, from a DIY wedding to virtual bonfire. These were all special experiences and I remember them dearly.

Finally, I would like to acknowledge Samuel Oliver Ay-Bryson who has filled my life with endless gratitude and unconditional happiness – thank you for having been by my side throughout my whole academic journey from bachelor to dissertation and beyond.



The authenticity of simulated patients in psychotherapy training and research

## Abstract

Mental health problems are highly prevalent worldwide. Fortunately, psychotherapy has proven highly effective in the treatment of a number of mental health issues, such as depression and anxiety disorders. In contrast, psychotherapy training as is practised currently cannot be considered evidence-based. Thus, there is much room for improvement. The integration of simulated patients (SPs) into psychotherapy training and research is on the rise. SPs originate from the medical education and have, in a number of studies, been demonstrated to contribute to effective learning environments. Nevertheless, there has been voiced criticism regarding the authenticity of SP portrayals, but few studies have examined this to date.

Based on these considerations, this dissertation explores SPs' authenticity while portraying a mental disorder, depression. Altogether, the present cumulative dissertation consists of three empirical papers. At the time of printing, *Paper I* and *Paper III* have been accepted for publication, and *Paper II* is under review after a minor revision.

First, *Paper I* develops and validates an observer-based rating-scale to assess SP authenticity in psychotherapeutic contexts. Based on the preliminary findings, it can be concluded that the Authenticity of Patient Demonstrations scale is a reliable and valid tool that can be used for recruiting, training, and evaluating the authenticity of SPs.

Second, *Paper II* tests whether student SPs are perceived as more authentic after they receive an in-depth role-script compared to those SPs who only receive basic information on the patient case. To test this assumption, a randomised controlled study design was implemented and the hypothesis could be confirmed. As a consequence, when engaging SPs, an in-depth role-script with details, e.g. on nonverbal behaviour and feelings of the patient, should be provided.

Third, *Paper III* demonstrates that psychotherapy trainees cannot distinguish between trained SPs and real patients and therefore suggests that, with proper training, SPs are a promising training method for psychotherapy.

Altogether, the dissertation shows that SPs can be trained to portray a depressive patient authentically and thus delivers promising evidence for the further dissemination of SPs.

## Zusammenfassung

Psychotherapie gilt als hoch wirksame Behandlung von psychischen Störungen, wie Depressionen oder Angststörungen. Trotz verpflichtender Psychotherapieausbildung zur Befähigung psychotherapeutischer Behandlungen gibt es wenig Evidenz, welche Methoden dem effektiven Training von angehenden Psychotherapeut\*innen dienen. Eine Lösung besteht im Einsatz von Simulationspatient\*innen (SPs) in der Psychotherapieausbildung sowie -forschung als nachweislich effektive Lehrmethode. Dabei werden jedoch SPs für ihren Mangel an Authentizität kritisiert, wobei ungenügend Studien diesen Aspekt direkt untersuchten.

Daher war es Ziel der vorliegenden Dissertation, die Authentizität von SPs, während sie eine psychische Störung (d.h. Depression) simulieren, zu untersuchen. Die vorliegende kumulative Dissertation basiert auf drei empirischen Arbeiten. Zum Zeitpunkt der Einreichung der Dissertation wurden *Paper I* und *Paper III* nach erfolgreichem Peer-Review in Fachzeitschriften zur Veröffentlichung angenommen, *Paper II* befindet sich nach einer Minor Revision im Peer-Review.

In *Paper I* wurde die Skala Authentizität von Patientendarstellungen zur Erfassung von Authentizität von SPs, die psychische Störungen darstellen, entwickelt und validiert. Die Ratingskala weist gute psychometrische Gütekriterien auf und ist für den Einsatz in Forschung und Praxis geeignet.

In *Paper II* wurde mittels einer randomisiert kontrollierten Studie demonstriert, dass die Ausarbeitung von Rollenanleitungen von SPs für deren Darstellung relevant ist: Studierende, die als SPs fungierten, wurden authentischer wahrgenommen, wenn sie eine detaillierte Rollenanleitung erhielten als jene SPs, die eine einfache Rollenanleitung erhielten.

In *Paper III* konnte gezeigt werden, dass Psychotherapeut\*innen in Ausbildung reale Patient\*innen von trainierten SPs nicht unterscheiden konnten. Der Einsatz von SPs ist demzufolge eine vielversprechende Trainingsmethode der Psychotherapie.

Insgesamt stellt die vorliegende Dissertation dar, dass SPs trainiert werden können, Patient\*innen mit Depressionen authentisch darzustellen. Die Arbeit liefert erfolgsversprechende Ergebnisse für die weitere Dissemination von SPs im Ausbildungskontext der Psychotherapie.

## List of original papers

***Paper I: “Development of the Authenticity of Patient Demonstrations scale”***

Ay-Bryson, D. S., Weck, F., Kühne, F. (2020). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000349>

***Paper II: “Enhancing SPs’ authenticity”***

Ay-Bryson, D. S., Weck, F., Kühne, F. (2021). Can students in simulation portray a psychotherapy patient authentically with a detailed role-script? Results of a randomized-controlled study. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000388>

***Paper III: “Indistinguishability of SPs from real patients”***

Ay-Bryson, D. S., Weck, F., Heinze, P. E., Lang, T., Kühne, F. (in press). Can psychotherapy trainees distinguish standardized patients from real patients? Results of a pilot study. *Zeitschrift für Klinische Psychologie und Psychotherapie*. <https://doi.org/10.1026/1616-3443/a000594>



# Table of contents

<b>1. Introduction .....</b>	<b>11</b>
<b>2. Theoretical and empirical foundations .....</b>	<b>13</b>
2.1 Psychotherapy .....	13
2.1.1 Efficacy and limitations of psychotherapy .....	13
2.1.2 The role of psychotherapists.....	14
2.2 Therapist training towards evidence-based training .....	16
2.2.1 Assessment of therapist competence .....	17
2.2.2 Therapist training in Germany.....	18
2.3 Summary and the relevance of simulated patients.....	19
2.4 Simulated patients.....	20
2.4.1 Definition, settings and methodology.....	20
2.4.2 Theoretical foundation of learning with SPs .....	22
2.4.3 Training effects on trainees .....	24
2.4.4 Benefits and drawbacks .....	25
2.4.5 A theoretical perspective on authenticity .....	27
<b>3. Aims of dissertation.....</b>	<b>29</b>
<b>4. Overview of papers.....</b>	<b>30</b>
4.1 Authenticity of Patient Demonstrations scale ( <i>Paper I</i> ).....	30
4.2 Enhancing SPs' authenticity ( <i>Paper II</i> ) .....	32
4.3 Indistinguishability of SPs from real patients ( <i>Paper III</i> ) .....	34
<b>5. Discussion.....</b>	<b>37</b>
5.1 Summary of dissertation contributions .....	37
5.2 Limitations and future research directions.....	43
5.2.1 Student SPs.....	43
5.2.2 Rater perspective and training .....	44
5.2.3 Portrayal of depression only .....	45
5.2.4 Lack of SP distress monitoring.....	45

5.2.5	Lack of relation between authenticity and learning effectiveness.....	46
5.3	General discussion .....	47
5.4	SPs and evidence-based training.....	48
5.5	Conclusion .....	48
<b>6.</b>	<b>References .....</b>	<b>50</b>
<b>7.</b>	<b>Appendix .....</b>	<b>60</b>
7.1	Paper I.....	60
7.2	Paper II.....	89
7.3	Paper III .....	120
7.4	Eidesstattliche Erklärung .....	149

## 1. Introduction

“Everything I’ve ever done (in a film) . . . , it requires this getting to some sort of emotional reality that is contrary to the actual setting that you’re in.” – Tom Hanks

Psychotherapy has been demonstrated to be highly effective in the treatment of a number of mental disorders (Lambert, 2013). However, psychotherapy training is evaluated less systematically and thus cannot be considered to be evidence-based (Fairburn & Cooper, 2011; Rakovshik & McManus, 2010). It can be argued that psychotherapists<sup>1</sup> need effective education in order to deliver evidence-based treatment properly. One promising approach to effective psychotherapy training is the simulated patient (SP) methodology. SPs are healthy people who simulate a patient with a disorder for training and assessment purposes (Barrows, 1993), which also allows their incorporation into research. In a similar way to Tom Hanks in his quote above, SPs are not suffering from a disorder in reality, but pretend to be. There are numerous benefits for SPs’ use in training and research, such as their standardisation and repeatability, which will be elaborated in more detail later. Further, most recently, the psychotherapy law in Germany was reformed (Entwurf Eines Gesetzes Zur Reform Der Psychotherapeutenausbildung, 2019). Accordingly, the law now prescribes university examinations of psychotherapy students integrating SPs, similar to the Objective Structure Clinical Examination (OSCE) in medical education. Hence, one can expect SPs to be integrated even more into psychotherapy education in the foreseeable future.

Nevertheless, there is also criticism regarding the use of SPs. In a recent scoping review on barriers and facilitators of SPs in clinical psychology and psychotherapy (Kühne et al., 2018), we found that many publications referred to authenticity of SPs portraying a disorder as a major concern. Specifically, authors noted that a gap or difference between real patients and SPs might result in difficulties in empathy or assessment. On the other hand, previous studies have demonstrated that it is possible for SPs to portray mental health problems authentically (Partschefeld, 2013; Wünderich et al., 2008). However, more empirical data on the question as to whether and how SPs can be

---

<sup>1</sup> The terms psychotherapist and therapist, psychotherapy training and training are used synonymously in the present dissertation.

trained to authentically portray a patient, is sparse. As a result, it was my aim to empirically study SP authenticity.

In other words, the present dissertation intends to contribute to research conducted on SPs with a view to improving this highly promising teaching method for psychotherapy training. Based on three studies, I intended to answer three questions: first, how to assess the authenticity of SPs by developing and validating an observer-based rating-scale; second, how to improve authenticity using a randomised controlled study design; and finally, whether psychotherapy trainees are able to distinguish between trained SPs and real patients using a within-between subject study design.

In the next section, I will first outline the theoretical as well as empirical background of my dissertation. Subsequently, I will provide an overview of the above-mentioned studies upon which my dissertation is based, followed by discussing the limitations of the present work and posing future research directions.

## 2. Theoretical and empirical foundations

In order to understand the need to train SPs and to study this approach as a training method for psychotherapists, it is useful to have some background information on psychotherapy in general and the training of psychotherapists. The first part of the theoretical and empirical background will focus on those. From there, I will focus on the SP methodology and specifically authenticity, i.e. the main focus of the dissertation.

### 2.1 Psychotherapy

#### 2.1.1 *Efficacy and limitations of psychotherapy*

Mental health problems are highly prevalent (Jacobi et al., 2014). In 2017, a study estimated that, worldwide, 792 million (10.7%) people were suffering from a mental health disorder (Ritchie & Roser, 2018). Fortunately, psychotherapy has repeatedly been demonstrated to be highly effective in the treatment of a number of mental health problems, such as mood and anxiety disorders: “The consistent finding of positive psychotherapy effects – across decades, thousands of studies and hundreds of meta-analyses, examining diverse disorders and therapies – is seemingly undebatable at this point in time” (Lambert, 2013, p. 176).

The majority of treated patients suffering from a mental health problem benefits from psychotherapy, i.e. experiences a symptom reduction. On the other hand, there remain patients who do not respond to treatment and thus do not improve despite undergoing treatment (Lambert, 2013). The literature even provides indices for a deterioration of symptoms after treatment compared to the beginning of treatment. Such an unfavourable course is referred to as a negative effect. According to Lambert (2013), 5% to 10% of adult patients experience increased symptoms during the period of psychotherapy. Consequently, Lambert (2013) concludes that “there is still plenty of room for improvement” (p. 191) in psychotherapy.

Needless to say, negative effects should be reduced, if not completely prevented. Therefore, one might ask what factors influence the occurrence of negative effects. In knowing this, we may come one step closer to improving psychotherapy. The literature suggests that negative psychotherapy effects may, among other factors, such as patient characteristics or external factors, be associated with the psychotherapist. Since the

present dissertation focusses on psychotherapy training, I will only focus on this factor. One paper from Germany analysed data on patient ( $N = 81$ ) complaints about their psychotherapists (Lange et al., 2009). The authors found that a lack of empathy of the therapist towards the patient was the most common of complaints. According to the survey, the lack of empathy caused that the patients could not develop trust in their therapist. The second most common complaint, according to this study, was that therapists paid too little attention to the problems mentioned by the patients. Other studies that also examined the relation between complaint-related factors and patient outcomes found that a poor therapist-patient-relationship (Hoffmann et al., 2008), and negative communication skills towards patients, such as belittling, ignoring or neglecting (Castonguay et al., 2010), are linked to negative effects of psychotherapy. In short, psychotherapists seem to play an important role in the context of negative therapy outcomes.

Fortunately, psychotherapy trainees can be enlightened and trained in this regard. Lambert (2013) concluded, in order to adequately prevent potential harm to patients as a result of psychotherapy, that special attention should be given to “the selection of students for graduate study, the selection of clients for treatment, the suitability of specific procedures for some clients, and the selection, training, and monitoring of therapists” (p. 192).

Altogether, there is clear evidence that psychotherapy is very effective in the treatment of a number of mental health problems. At the same time, attention should be paid to possible negative effects. Specifically, psychotherapists may contribute to improving psychotherapy. Therefore, in the next section we will focus more closely on the role of psychotherapists and their contribution to patient outcomes.

### ***2.1.2 The role of psychotherapists***

In contrast to well documented positive effects of psychotherapy on a wide range of mental health problems, research on therapist effects has been long neglected (Barkham et al., 2017; Johns et al., 2019). Fortunately, recent psychotherapy research has witnessed a crucial step towards therapist-focussed-research, which, as the term implies, focusses on therapist effects. Therapist effects can be understood as “the contribution that can be attributed to therapists when evaluating the efficacy of a psychological intervention” (Lutz & Barkham, 2015, p. 1, as cited in Barkham et al., 2017).

In a prominent large-scale study, Okiishi et al. (2006) examined differences between individual therapists over the course of six years. Client participants of this study were college students and the most common diagnoses were mood, anxiety and adjustment disorders. The results of this study indicated that the type and amount of therapist training, theoretical orientation, and gender of the therapist did not significantly contribute to the speed of patient improvement. However, therapists did significantly differ from one another with regard to the number of sessions for which they had seen their clients, and the overall patient improvement based on pre- and post-therapy change scores. In some cases, there were even strong differences, “for example, the rate of deterioration of clients seen by the bottom-ranked therapists was double that found in clients of the top-ranked therapists” (p.1167).

The significant variation between therapists regarding their clients’ rates of improvement was confirmed in a later literature summary (Baldwin & Imel, 2013). The authors also concluded that psychotherapists vary and hence play a crucial role in patient outcomes. The results once more highlighted the importance of therapist-focused research, for which reason the authors urged for more therapist effect studies. In a more recent systematic review, Johns and colleagues (2019) examined the literature for therapist effects since Baldwin and Imel’s (2013) review. In their paper, the authors confirmed (a) previous evidence that a difference in therapist effectiveness can be found across clinical settings, patient groups as well as across routine practice or trial data; and (b) that therapists, as was suggested previously by Baldwin and Imel (2013), contribute to the patient outcome variability.

The general consensus across studies is that therapist effect sizes range from 5% to 8% depending on the study (Baldwin & Imel, 2013; Barkham et al., 2017). This effect size is comparable with one of the strongest therapy outcome predictors, such as therapeutic alliance (Baldwin & Imel, 2013), for which reason therapist effects should be taken seriously. Given these findings, it should be in the interest of psychotherapy that therapists be able to provide effective therapy for all patients. One possible way to enable psychotherapists to do so might be effective training methods. Otherwise, “if we cannot show that therapists become more expert as a result of training and practice, our current models of graduate training, as well as continuing postdoctoral education, need to be re-examined” (Hill et al., 2017, p. 8). Accordingly, training becomes of crucial importance

as it may have a great effect on therapists in training, which in turn may be linked with patient outcome. For this reason, in the next chapter therapist training will be highlighted.

## **2.2 Therapist training towards evidence-based training**

Psychologists undergo an extensive psychotherapy training that usually takes up to several years prior to becoming a licensed psychotherapist. Training structure, modules and lengths may differ dependent on the theoretical orientation one pursues (e.g. cognitive-behaviour therapy or psychoanalysis). Regardless of the theoretical orientation, however, training might be seen as an important period of time for future psychotherapists. All in all, the primary goal of psychotherapy training is to equip therapists-in-training with an armamentarium of skills that they can flexibly use in practice in response to their clients' needs (Hill & Knox, 2013). Thus, training is a fundamental period during which helping or therapeutic skills are acquired.

Nevertheless, in contrast to numerous empirical studies on the efficacy and effectiveness of psychotherapy as was outlined above, psychotherapy training can yet not be considered evidence-based as research on training methods is still in its infancy (Callahan & Watkins, 2018; Fairburn & Cooper, 2011; Rakovshik & McManus, 2010). This may be one reason why training methods have changed little over time (Fairburn & Cooper, 2011).

Some evidence is already available on the effectiveness of various teaching methods. Hill and Lent (2006), for instance, found in their meta-analysis that the following specific methods were significantly more effective than no training in teaching basic helping skills: instruction ( $d = .63$ ), model learning ( $d = .90$ ), and feedback ( $d = .89$ ). Moreover, Hill and Knox (2013) cited an investigation on the teaching of empathic communication (Nerdrum & Rønnestad, 2002), in which the authors found that the most valued training methods reported by trainees was theory, role playing and model learning. Nevertheless, Hill and Knox (2013) cautioned against overreliance on the reported results as most of the studies exhibit methodological problems, such as a lack of clear definitions, random assignments and control conditions. Hence, further robust research is needed to conclude which training tools are effective, in order to determine which approaches may have a positive effect on trainees' therapeutic skills, anxieties and self-efficacy (Hill & Knox, 2013).

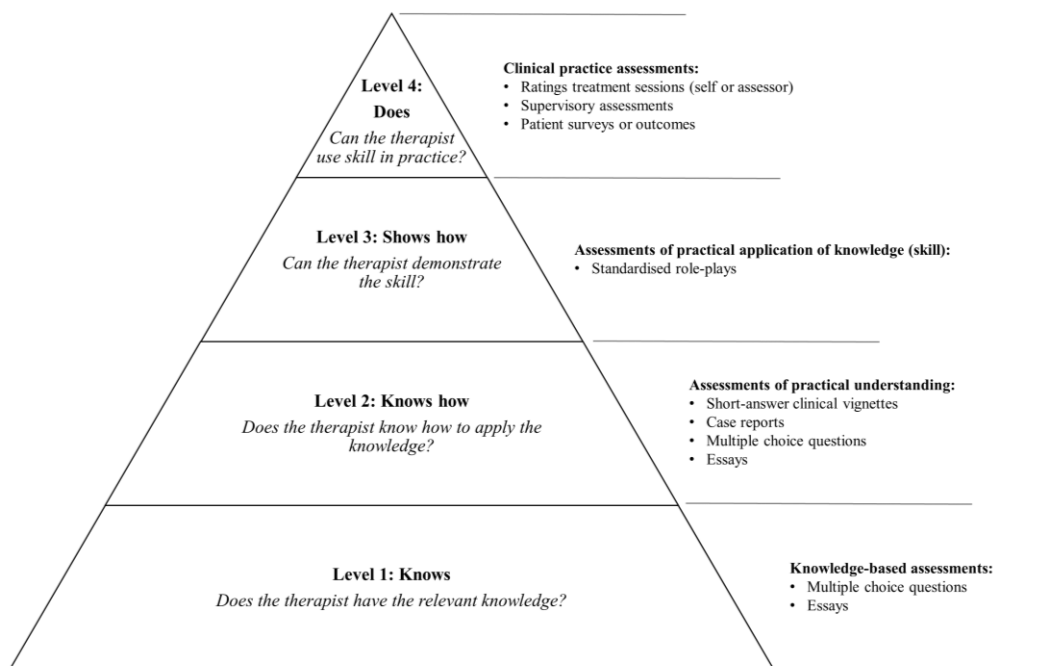


There are various ways of examining the effectiveness of training. For example, qualitative studies can be helpful in gaining insight into trainees' perspectives. This way, perceived challenges and useful training approaches may be identified (e.g. Hill et al., 2007; Nerdrum & Rønnestad, 2002). Another possibility is to conduct quantitative research. McHugh and Barlow (2010) suggested incorporating the assessment of training outcomes and clinical outcomes into training programmes, whereby objective training outcomes contain therapist competence. Indeed, psychotherapeutic competence is highly relevant to the evaluation of psychotherapy training methods (Kühne et al., 2020). Therefore, in the following section I will illustrate a much-cited framework on the assessment of clinical skills.

### ***2.2.1 Assessment of therapist competence***

Ultimately, one goal of psychotherapy training is to train trainees to become competent therapists. Hence, therapeutic competence becomes an important factor. In his review, Miller (1990) proposed a framework for the assessment of clinical skills, competence and performance. This framework was later extended and adapted to the context of psychotherapeutic competence by Muse and McManus (2013; see Figure 1). On this pyramid model four levels are defined, namely *knows*, *knows how*, *shows how* and *does* (Miller, 1990; Muse & McManus, 2013). Each of the levels can be assessed in different ways. The most basic level of knowledge (*knows*) can be assessed by essays or multiple choice questions. For the second level of practical understanding (*knows how*), additionally short answer clinical vignettes and case reports can be deployed. For the assessment of the third level, i.e. practical application of knowledge (*shows how*), Muse and McManus (2013) suggested standardised role-plays. Finally, the fourth level, namely the clinical practice (*does*), can be assessed by rating treatment sessions, through supervisory assessments or patient surveys or outcomes. Altogether, this framework illustrates the stages of developing clinical competence from a very basic to an advanced and practical level.

Often, university graduation is followed by a postgraduate psychotherapy training. Usually, trainees learn and acquire the necessary skills for therapeutic work throughout their training. In the following, therefore, therapist training as regulated in Germany will be briefly outlined.



**Figure 1.** Framework of cognitive behaviour therapist competence assessments, retrieved from Muse & McManus (2013, p. 487), based on Miller's (1990) clinical skill hierarchy.

### 2.2.2 Therapist training in Germany

Psychotherapy training in Germany comprises a minimum of 4200 hours which usually takes up to three to five years and is regulated by the local psychotherapy law ("Psychotherapeutengesetz") and regulation ("Ausbildungs- und Prüfungsverordnung für Psychologische Psychotherapeuten"). The training program is structured upon the following modules:

- 600 hours of theoretical training ("Theoretische Ausbildung")
- 1200 hours at a psychiatry ("Praktische Tätigkeit 1")
- 600 hours at a psychosomatic facility ("Praktische Tätigkeit 2")
- 600 hours of independent treatment of outpatients accompanied by a minimum of 150 hours of regular supervision ("Praktische Ausbildung und Supervision")
- 120 hours of self-practice ("Selbsterfahrung")
- 930 hours of free choice ("Freie Spitze")

Since the present dissertation focusses on SPs, which can be allocated to the theoretical module of psychotherapy training, I will take a closer look at this stage. The theoretical

training contains a minimum of 600 hours, of which 400 hours are dedicated to advanced knowledge in the chosen theoretical orientation and 200 hours are dedicated to basic knowledge of alternative therapy approaches. For example, when a trainee undergoes training in behaviour therapy, 400 hours of theory are dedicated to behaviour therapy and 200 hours are dedicated to psychodynamic and other approaches. The theoretical training can be implemented in different ways, such as in form of a lecture, seminar and/or exercises, whereby lecture forms may not exceed a third of the total theoretical training (Partschefeld, 2013). The goal of the theory classes is to impart theoretical knowledge as well as practical therapeutic skills. Unfortunately, the transfer from theory into practice is not always successfully implemented, for which reason Strauß and Brähler (2009) urged for more practically oriented exercises and training units, highlighting the importance of this transfer.

Finally, most recently the psychotherapy law was reformed (Entwurf Eines Gesetzes Zur Reform Der Psychotherapeutenausbildung, 2019). The new law in Germany prescribes university examinations of psychotherapy students, similar to the OSCE in medical education. Hence, in the future, trainees will be assessed in a more structured and practical manner with regard to their clinical competence.

### **2.3 Summary and the relevance of simulated patients**

In the sections above the efficacy and limitations of psychotherapy were exemplified. While psychotherapy is evidently effective in the treatment of a number of mental health problems, a substantial percentage of patients still either do not respond to treatment or their symptoms even deteriorate over the course of treatment. Notably, some studies demonstrated that negative effects can be linked to the psychotherapist. Also, there is considerable evidence that psychotherapists differ greatly among each other. The training of psychotherapists is thus crucially important in the context of health care for patients suffering from mental health problems, yet little research has been conducted on the effectiveness of training methods. Although it can be concluded that training is more effective than no training, most of the studies have great methodological problems which limit our insights into training research and leave many questions unanswered. Thus, there is a need for more robust research in psychotherapy training. One very promising approach to address the current issues is the SP methodology.

At a glance, SPs possess the potential to contribute to:

- evidence-based psychotherapy training,
- the transfer from theory into practice, and
- the assessment of clinical competence, as was suggested by Muse and McManus (2013) and as will be required in the newly reformed psychotherapy training in Germany.

Altogether, it seems important to deliver more systematic research on SPs given their ever-increasing use. In the following sections, I will hence concentrate on the SP methodology, which then lead to the aims of the dissertation.

## **2.4 Simulated patients**

### **2.4.1 Definition, settings and methodology**

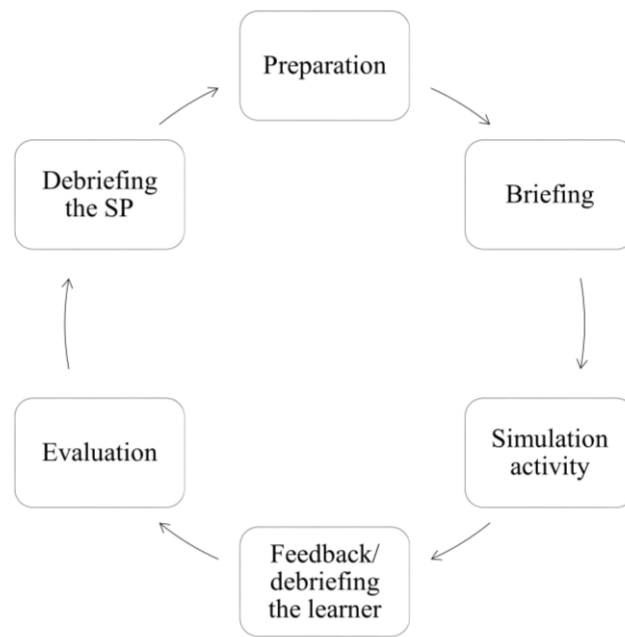
There exists a variety of terms, such as programmed, prepared, trained, volunteer or actor patient, role-player, patient instructor or incognito or unannounced patient. All of these terms refer to similar concepts of simulated or standardised patients (Kühne et al., 2018; Nestel & Bearman, 2015). Since the focus of the current dissertation is simulated and standardised patients, we will concentrate on these terms. While SPs are defined as healthy laypersons who simulate a disorder with respective symptoms and personality traits for teaching purposes, the term “standardised patients” refers to the same definition with the addition that they provide the same role-play repeatedly, and in a consistent manner (Adamo, 2003; Barrows, 1993).

Since their first introduction in the 1960s in the United States of America in the context of clinical neurology (Barrows & Abrahamson, 1964), SPs have increasingly been deployed in Europe mostly at medical faculties as well as worldwide in various professional settings. In addition to their prevalent implementation in medical training, SPs contribute to a wide range of professional education, such as in nursing, pharmacy, or physiotherapy (Nestel et al., 2015). In recent years, SPs have also been integrated in psychotherapy training (Partschefeld et al., 2013). Across all professions, SPs can equally contribute to training, assessment and research.

For instance, in nursing training, SPs were incorporated into an educational intervention with the aim to improve advanced practice nurses’ communication skills in the context of “breaking bad news” to cancer patients (Eid et al., 2009). Similarly, in

physiotherapy training, SPs were implemented in order to promote the students' interpersonal skills (Lewis et al., 2008). Possibly the most prominent example for the use of SPs in assessment is the Objective Structured Clinical Examination (e.g. Harden et al., 1975; Newble, 2004), which examines a medical student's clinical competences through an SP-based demonstration. Finally, SPs are increasingly being used in psychotherapy training research. In an ongoing research project SPs are implemented to evaluate and compare different training approaches with students to develop competences (Kühne et al., 2020).

There are six established phases in SP-based education (Figure 2) that are considered to create effective training experiences (Nestel & Bearman, 2015). *Preparation* refers to all stages prior to the action of simulation-based education, such as recruiting and training SPs, setting learning goals and designing scenarios. During the *briefing* phase, the simulation-based process is explained to everyone involved. This phase may also be the time during which trainees share prior experiences or SPs could be checked if they know their role well. The next phase is the *simulation activity*, which is where the simulated interaction between trainee and SP takes place. This is usually the phase that varies the most, depending on which educational context and for what purpose, such as assessment versus learning modality, it takes place. This phase is then followed by the *debriefing and feedback* phase, during which the trainees' feelings are checked, feedback is provided and during which ideas for future implementation can be exchanged. Finally, during the *evaluation* phase successes as well as limitations of the session are summarised. There is an optional addition, namely the phase of *debriefing the SP*. It is understood as an opportunity to assist the SPs to step out of their roles, which is sometimes referred to as "de-roling". This phase may especially be helpful if SPs have had to simulate an emotionally challenging role.



**Figure 2.** Simulated patient methodology, retrieved from Nestel & Bearman (2015).

Those who are unfamiliar with SPs, however, may wonder how trainees can learn with SPs during their training. In order to address this question, an underlying theoretical background on two learning theories will be given next.

#### **2.4.2 Theoretical foundation of learning with SPs**

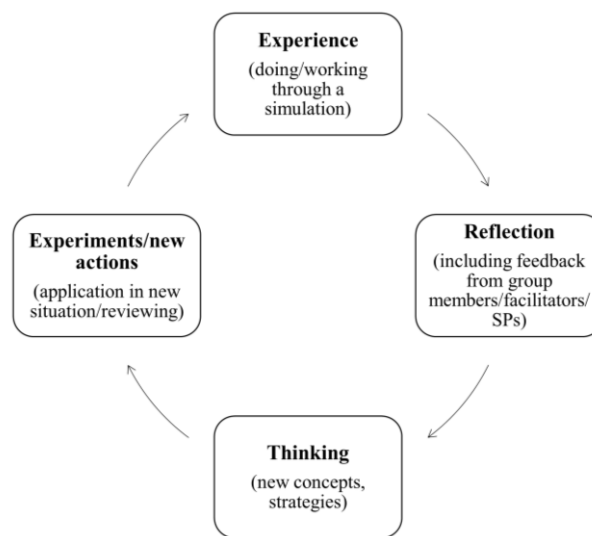
Arguably, “the best way to develop skills is to consolidate learning through practice” (Thistlethwaite & Ridgway, 2015, p. 18). According to Thistlethwaite and Ridgway (2015), training with SPs or simulation-based training enables a trainee to learn by skill-practising and to learn by experiencing. Two theory-based approaches to learning, namely Kolb’s experiential learning cycle (Kolb et al., 2001) and the cognitive load theory and scaffolding (Bearman & Nestel, 2015), have been highlighted in the context of simulation-based training since both theories emphasise learning through practice.

##### *Kolb’s experiential learning theory*

Kolb’s experiential learning theory (2001) understands learning as learning from experience. Further, it understands learning as a process including feedback on the effectiveness of learning efforts. It includes experience, reflection, thinking and experiments and new actions (see Figure 3). Incorporating the experiential cycle to simulation-based education may enhance trainee engagement during simulations (Abdool

et al., 2017). Since the trainee profits from the advantage that the simulated environment is structured and controlled (Thistlethwaite & Ridgway, 2015), he or she can reflect upon the experience including observations and feedback provided by a supervisor, group members, the SP or other peers. Based on the reflections made, the trainee can then think about new concepts and strategies and can apply these in a new (simulated) situation.

Thistlethwaite and Ridgway (2015) underline that the degree of structure may differ between trainees depending on their prior experience with role-play and/or simulation-based learning and interactive learning in a group including feedback. As a guideline, they summarise that generally more inexperienced trainees, such as novice trainees, tend to require a higher degree of structured activities compared to more experienced trainees. In short, according to Kolb's theory, experiential learning in SP methodology is reached through the practical understanding of trainee-SP interactions and the environment that surrounds it.



**Figure 3.** Kolb's learning cycle applied to simulation activity, retrieved from Thistlethwaite & Ridgway (2015).

### *Cognitive load theory and scaffolding*

The cognitive load theory is an empirical learning theory that primarily focuses on the working memory of a learner (Bearman & Nestel, 2015). Specifically, it acknowledges that the working memory is highly limited. As Van Merriënboer and Sweller (2010) noted, the working memory cannot actively process more than two to four elements and

its time capacity for processing is highly restricted. Consequently, extraneous (i.e. irrelevant) load needs to be decreased in order to increase germane (i.e. relevant) load. While irrelevant load may, for example, be reduced by goal-free tasks or completion tasks, relevant load may be optimised by increasing variability of the tasks or by evoking self-explanation (Van Merriënboer & Sweller, 2010). For instance, during training, irrelevant load can be decreased by focusing on smaller goals, such as practising behaviour activation with the SP, in order to increase learning effects.

As a complement to the cognitive load theory, Bearman and Nestel (2015) suggested scaffolding, which is a means of supporting the trainees and can be especially beneficial when a trainee is experiencing difficulties with limited capacity in the working memory. The core understanding of scaffolding is that scaffold support is removed slowly so that the trainee can apply an acquired skill independently (Pea, 2004). Scaffolding as applied to the SP methodology is for instance that “in a breaking bad news scenario, novices can be refocused and guided by the SP, but as the student skill and confidence progress, the SP may provide fewer cues.” (Bearman & Nestel, 2015, p. 35).

### ***2.4.3 Training effects on trainees***

Taking into account that SPs provide the opportunity to learn experience-based and also potentially provide support during training, it is not surprising that several studies and meta-analyses have demonstrated that SPs contribute significantly to learning effects across health profession education, mostly in medicine and nursing (e.g. Cook et al., 2011, 2012; Eckel et al., 2014; May et al., 2009; McGaghie et al., 2011; Piot et al., 2020; Zendejas et al., 2013). Specifically, learning effects have been found regarding clinical knowledge and skills, self-confidence and comfort level during communications, empathy and communication skills. This implies that trainees who were trained with SPs reported changes in knowledge, skills or attitudes toward the patient as well as increased self-confidence during communications (e.g. May et al., 2009). Similarly, SPs seem to contribute to an increased capability for empathy in the trainees as well as an interest in the subject they are trained in due to the role-plays with SPs (e.g. Eckel et al., 2014). Interestingly, one systematic review on simulation-based education in medicine (Zendejas et al., 2013) reported that simulation-based education was linked to improved patient outcomes with small to moderate effects. All in all, there is good reason to assume that the SP method is effective for training in health professions.



One study (Partschefeld et al., 2013) specifically examined whether the SP method results in increased therapeutic skills, empathy, therapeutic alliance and self-efficacy in the context of psychotherapy training. The authors of this study could demonstrate that psychotherapy trainees who were trained with SPs significantly increased their scores on all of these variables. Therefore, the authors concluded that the SP training method is effective and may contribute significantly to the acquisition of therapeutic skills, even though it is costly and time-consuming. Finally, Partschefeld et al. (2013) underlined that SPs possess numerous advantages, which I will summarise in the following section.

#### **2.4.4 *Benefits and drawbacks***

Even though higher external validity can be expected with real patients, training with real patients exhibits a number of problems compared to SPs (Kühne et al., 2020). Problems include the dependence of competence assessments on the patient difficulty and their behaviours. Further, several measurements are required for reliable assessments of competence (Dennhag et al., 2012). Patients may also refuse to participate in training settings. In contrast, SPs have a variety of benefits (see Table 1) over real patients in the context of training. At the same time, they also have drawbacks that need to be considered. In the following, the advantages as well as disadvantages of SPs will be depicted.

To begin with, SPs are standardised, thereby enabling comparability between situations. Hence, in examinations, such as the OSCE, students can be examined in comparably difficult situations. Further, SPs can be employed repeatedly. Consequently, research or training can be implemented based on various time points, which enables more reliable measurements. Moreover, SPs' availability is a major benefit for research as well as training. SPs can volunteer in established large-scale data bases, such as at the SP programme at the Charité clinic (Charité Universitätsmedizin Berlin, 2020), where medicine students are trained. This way, SPs are more readily available for participation in a simulation, and also for different portrayals (Eagles et al., 2007). In this context, the complexity of interactions can be controlled in adaptation to trainees' current levels and needs (Pheister et al., 2016). Furthermore, through the implementation of SPs, trainees are given the opportunity to learn through various learning mechanisms, such as experience-based learning, model learning and learning through individualised feedback provided by the SPs (Weaver & Erby, 2012). Regarding the latter, asking real patients for

feedback after a treatment session may not always be appropriate, for which reason SPs are more viable for providing feedback. Similarly, the use of SPs enables the opportunity to use recordings of audio or videos. According to McMahon and Ledden (2019), such recordings possess a number of advantages, such as repeated and holistic playing of a session that may promote the acquisition of competence. Further, audio and videos can be used even long after the actual experience (Pheister et al., 2016). Moreover, the simulated environment can provide trainees with more confidence when encountering an anxiety-provoking situation (Pheister et al., 2016), and provides reassurance to the trainees in emotionally challenging situations. In this vein, the use of SPs avoids potential mistreatment of real patients (Eagles et al., 2007). SPs also enable more predictable training experiences for crucial clinical issues (Pheister et al., 2016) and less common situations, such as domestic violence and emotionally difficult patients (Eagles et al., 2007). Finally, simulated encounters can be frozen in order to give corrective feedback, which would be impossible or at the very least problematic with real patients (Eagles et al., 2007).

**Table 1.** Overview of benefits and drawbacks of simulated patients.

<b>Benefits</b>	<b>Drawbacks</b>
<ul style="list-style-type: none"> <li>• Standardised portrayal</li> <li>• Repeatability</li> <li>• Availability</li> <li>• Adaptability/Flexibility</li> <li>• Provision of individualised feedback</li> <li>• Audio and video recordings</li> <li>• Safe learning environment</li> <li>• Reassurance during emotionally challenging situations</li> <li>• Less common and more difficult clinical situations</li> <li>• Simulations can be frozen</li> </ul>	<ul style="list-style-type: none"> <li>• Cost implications</li> <li>• Require staff and facilities</li> <li>• Require organisation time</li> <li>• Require technology</li> <li>• Lack of realness</li> <li>• SPs may experience adverse symptoms</li> </ul>

Despite SPs' many benefits as shown above, they are also associated with some disadvantages that are worth considering. SP programmes are fairly costly as they require a lot of resources, such as money, (trained and knowledgeable) staff, facilities, organisation time as well as appropriate technology. On the other hand, once an SP

programme is established, it can continuously be used or role-scripts can be easily adapted. Further, since SPs are not real patients, many concerns have been voiced with regard to external validity, i.e. a potential lack of authenticity. One criticism towards SPs is that their portrayal is more similar to a textbook than a real-life situation (Wallace et al., 2002). Lastly, SPs can actually be stressed by the roles they demonstrate as they may experience adverse symptoms (Bokken et al., 2004). However, de-roling techniques and debriefing after the simulation may help SPs prevent such induction (Kühne et al., 2018).

#### **2.4.5 A theoretical perspective on authenticity**

Even though *actor patient* is a commonly used synonym referring to SPs, it does not necessarily imply that the actor patient is an actor. While there are similarities between acting performance and SPs, there are also fundamental differences (Smith, Edlington et al., 2015). Smith et al. (2015) importantly point to the difference in the purpose of performance. While an actor pretends to be someone else, i.e. a fictive character, as a “service of the playwright, a director’s vision or a producer” (p. 40), the SP has the purpose to portray a patient, through which he or she acts as a teacher; “ultimately, for SPs, it is always about the learner.” (p. 40).

Arguably, in order to enable trainees to experience real-life encounters during their training, SPs are required to be authentic. In a review recently conducted (Kühne et al., 2018) on facilitator and barrier factors of SPs for implementation in clinical psychology and psychotherapy, it became evident that many authors working with SPs raised concerns regarding SPs’ lack of authenticity. Authenticity can be defined as the “impossibility of distinguishing SPs from (real) patients” (Wünderich et al., 2012, p. 501). Thus, ideally, trainees should not be able to tell an SP from a real patient and vice versa. Authenticity of SPs’ portrayal is regarded as important since a lack of authenticity may dampen learning effects of trainees (Murtagh, 2015). For instance, one study (Krahn et al., 2002) showed that SPs were perceived as less authentic compared to real patients and that this led to difficulties in feeling empathetic towards the SPs.

Russian theatre practitioner Konstantin Stanislavski (1863 – 1938) explored ways to support the actor find into a role to portray characters authentically. According to Stanislavski, training is a process to connect cognitive, physiological and psychological domains with a view to creating an authentic portrayal.

Based on Stanislavski's approach, Nestel, Fleishman and Bearman (2015) proposed an adaptable approach to training SPs, that can be summarised into four training stages: (1) developing the character, (2) explaining the learning goal to the SP, (3) exploring the clinical setting, and (4) rehearsing. The overall goal is that the "character of the person to be portrayed remains prominent, allowing the group of SPs to develop a shared and coherent understanding of the SP role, the scenario and the overall activity" (Nestel et al., 2015, p. 63). Finally, it has been noted that not all SPs are equally suitable and capable of portraying all characters regardless of the time spent in training and preparation; for instance, more complex and challenging characters may require more experienced SPs (Smith et al., 2015).

Despite some theoretical understanding of authenticity, little empirical data exists currently. Studies incorporating SPs in their studies, have each different approaches to training and instructing SPs, which complicates the comparison between studies. Consequently, there exists no consensus or standard on SP authenticity and on how to train them. The heterogeneity of SP trainings and of the role-scripts used is described in more detailed in *Paper II*. Empirically, authenticity has seldom been the main subject of previous studies and the results are somewhat contradicting. Few studies have directly examined authenticity of SPs portraying a mental disorder (Fussell et al., 2009; Partschefeld et al., 2013, 2013; Wüdrich et al., 2012; Wüdrich et al., 2008). Furthermore, few studies have examined the relation between SP training and authenticity directly; one study that did (Perera et al., 2009) focused on medical education of haematology. SPs were rarely compared with real patients and those studies that did reported inconsistent findings (Krahn et al., 2002; Wüdrich et al., 2012). Therefore, our knowledge on the authenticity of SPs portraying a mental health problem remains preliminary. With the present dissertation I aimed to contribute to research on SPs taking into account limitations of previous studies. Subsequently, I will display the aims I pursued in the dissertation.

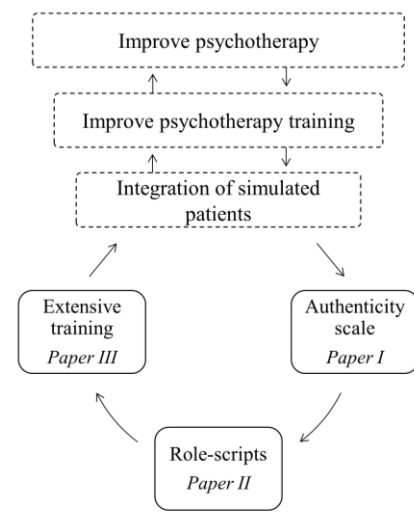
### 3. Aims of dissertation

As was argued in the previous sections, the use of SPs, due to their many benefits, bears the potential to improve training in psychotherapy. Furthermore, despite SPs' widespread use in medical education, where it has been common practice for decades, its implementation is only now starting to disseminate more into clinical psychology and psychotherapy (Kühne et al., 2018). However, evidence on the authenticity of SPs portraying a mental disorder is sparse. Therefore, the major interest of the present dissertation was to investigate SPs' authenticity in psychotherapeutic contexts. To this end, three studies were conducted. Altogether, the following aims were pursued:

- Aim 1:** Develop and validate an observer-based rating-scale that can be used to assess SPs' authenticity, specifically designed for psychotherapeutic settings (*Paper I: "Authenticity of patient demonstrations scale"*).
- Aim 2:** Study whether student SPs can portray a mental health problem more authentically after receiving an in-depth role-script compared to a baseline instruction (*Paper II: "Enhancing SPs' authenticity"*).
- Aim 3:** Investigate whether SPs can be distinguished from real patients by psychotherapy trainees following an exhaustive SP training (*Paper III: "Indistinguishability of SPs from real patients"*).

## 4. Overview of papers

The present dissertation is based on three original papers (Figure 4). In the following, each study will be depicted and major findings will be summarised. All manuscripts can be found in the Appendix. The presented papers explore different aspects related to the authenticity of SP portrayals. *Paper I* describes the development and validation of an observer-based rating instrument to assess SPs' authenticity. *Paper II* tests the hypothesis that a role-script with detailed information on a patient case increases authenticity in comparison to a role-script with little basic information. Lastly, *Paper III* explores the question as to whether psychotherapy trainees are able to distinguish SPs following an extensive training from real patients.



**Figure 4.** Structure of presented dissertation studies exploring the authenticity of simulated patients integrated in psychotherapy training. Dotted lines represent a visualisation of the theoretical background and are not subject of the present dissertation.

### 4.1 Authenticity of Patient Demonstrations scale (*Paper I*)

Ay-Bryson, D. S., Weck, F., Kühne, F. (2020). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000349>

### *Theoretical background*

A lack of authenticity of SP portrayals has repeatedly been reported as a barrier for the implementation of SPs (Kühne et al., 2018). On the other hand, a pilot-study demonstrated that it is possible for SPs to portray mental health problems authentically (e.g. Wünderich et al., 2008). Although many studies report on the authenticity of SPs, little to no empirical data is available, which prevents us from drawing firm conclusions on how to train SPs with regard to authenticity. Further, in order to investigate authenticity and potential influence factors that may be positively associated with authenticity, a reliable and valid instrument is necessary for the assessment. To date there exists no instrument for the examination of SPs' authenticity, especially designed for psychotherapeutic settings. Therefore, the main aim of this study was to develop and pilot-validate an observer-based rating instrument for the assessment of SP authenticity. The herein newly developed instrument was to be an easy-to-use rating tool that can be used regardless of what mental disorder is to be demonstrated.

### *Methods*

To develop the rating scale *Authenticity of Patient Demonstrations* (APD), we underwent a multi-step development and evaluation. The development included an inductive search for the most adequate items, followed by test phases and rewording of items as well as the implementation of an online expert survey with psychologists and theatre associates ( $N = 10$ ) with a view to content validity. Finally, the APD was translated into English applying the dual panel method (Hagell et al., 2010). The APD comprises ten items and uses a 4-point Likert scale from 0 = "strongly disagree" to 3 = "strongly agree". The evaluation of the APD was performed online using a within-subjects study design. A total of 97 participants ( $n = 85$  psychotherapy trainees;  $n = 12$  licensed psychotherapists) aged 24 to 48 years ( $M = 31.49$ ,  $SD = 5.17$ ) took part in this study as raters. They were presented two three-minute therapy segments with one featuring an authentic SP and one featuring an unauthentic SP based on a clinical depressive case. Raters were instructed to fill in the APD after each video, including further measures to perform analyses of reliability, validity and dimensionality of the APD. The order of the videos alternated.

### *Major findings*

Results of the content validity survey showed that most items were judged as comprehensible and relevant, whereas two items were perceived as less so. Items were adjusted accordingly. Results of the within-subjects online study demonstrated that the APD significantly differentiated between an authentic and an unauthentic SP ( $t(96) = 16.70, p < .001$ ; Cohen's  $d = 2.35$ ). Further, regarding reliability of the APD, we found Cronbach's  $\alpha$  to be .83, indicating good internal consistency (Kevala & Moosbrugger, 2012). Results of the exploratory factor analysis indicated a one-factorial structure accounting for 38% of the variance. Further, regarding convergent validity, a strong positive correlation between the APD and an instrument assessing SPs' performances within medical education, i.e. the Nijmegen Evaluation of the Simulated Patient (Bouter et al., 2013), was found ( $r = .82, p < .001$ ). Finally, we found an order effect of presentation, namely that raters judged the authentic SP as even more authentic ( $M = 2.65, SD = .34$ ) when they first saw the unauthentic SP than when the authentic SP was rated first ( $M = 2.35, SD = .38; t(93.28) = -4.19, p < .001$ ; Cohen's  $d = -.85$ ). Similarly, raters who first saw the unauthentic SP rated the person as even less authentic when raters first saw the authentic SP ( $M = 1.31, SD = .55$ ) in comparison to when the unauthentic SP was seen first ( $M = 1.54, SD = .46; t(91.60) = -2.28, p < .05$ ; Cohen's  $d = -.47$ ).

### *Conclusion*

This study demonstrated that SPs can authentically portray a depressive patient. The results of the present paper provide preliminary evidence for psychometrically sound properties of the APD and that the scale contributes to distinguishing between authentic and unauthentic SPs and may thus foster SPs' dissemination into evidence-based training. Further validation of the APD is required regarding a bigger sample size of SPs and interrater reliabilities, the comparison between SPs with real patients, and mental health problems other than depression.

## **4.2 Enhancing SPs' authenticity (*Paper II*)**

Ay-Bryson, D. S., Weck, F., Kühne, F. (minor revision). Can students in simulation portray a psychotherapy patient authentically with a detailed role-script? Results of a randomized-controlled study. *Training and Education in Professional Psychology*.



### *Theoretical background*

Currently there exists no scientific consensus on how to develop role-scripts for SPs (Davies et al., 2020). Some authors advocate role-scripts rich in details, whereas others argue for little information in the script, or even for instructing SPs to use personal problems for portrayal (Klaman & Yudkowsky, 2002; Kühne et al., 2018). The use of scripts in studies varies from structured scripts (Coyle et al., 1998), to no specific script (Imel et al., 2014). Taken together, there is much variance in the way of instructing SPs to demonstrate a patient with a mental disorder. This study, therefore, aimed to address the inconclusiveness regarding role-scripts and its relation with authenticity. Specifically, it was examined whether an in-depth role-script enhances SPs' authenticity compared to a role-script offering basic information.

### *Methods*

To test our hypothesis, a randomised controlled study design was used. Following results of an a priori power analysis that recommended a minimum sample of 54 participants, we recruited 60 student SPs, of which one participant dropped out after receiving the baseline instruction. The student SPs ( $N = 59$ ), recruited from the University of Potsdam, were randomly allocated to an experimental (i.e. in-depth role-script;  $n = 29$ ) or control condition (i.e. text task;  $n = 30$ ). Mean age of participants was 24.34 years ( $SD = 4.12$ ) and the majority (72.88%) was female. The SPs were asked to portray a depressive patient with a standardised study therapist (DAB) at two time points (pre and post). For simulation pre, all SPs received the same baseline information on a fictional depressive patient. For simulation post, SPs of the experimental group received an in-depth role-script with more detailed information in addition to the baseline information, whereas participants of the control condition received an unrelated text task. All simulations were video recorded. SPs' authenticity was evaluated from four perspectives using the APD scale: from the study therapist involved in the simulations (DAB), two independent student raters, an independent licensed psychotherapist and the SPs themselves after the second simulation. External raters, i.e. student raters and licensed psychotherapist, received a rater training on how to apply the APD. The rater training consisted of two parts that are considered effective training components for accurate assessment (Feldman et al., 2012).

### *Major findings*

We found significant interaction effects of time and condition based on evaluations from the external student raters' ( $F[1, 114] = 13.66, p < .001, \eta^2 = .09$ ) and the study therapist's (DAB) perspectives ( $F[1, 106] = 11.94, p < .001, \eta^2 = .08$ ). These results imply that there was a significant increase in authenticity at the second time point, and that this increase differed significantly between both conditions, in that SPs of the experimental group ( $M_{\text{students}} = 2.31, SD_{\text{students}} = .53; M_{\text{therapist}} = 2.46, SD_{\text{therapist}} = .46$ ) were rated as significantly more authentic than SPs of the control condition ( $M_{\text{students}} = 1.49, SD_{\text{students}} = .68; M_{\text{therapist}} = 1.50, SD_{\text{therapist}} = .72$ ). Since both other perspectives (licensed psychotherapist and SPs) only rated the authenticity based on the second simulation, an interaction effect could not be calculated, and is thus not included in this overview.

### *Conclusion*

This study demonstrated that in-depth role-script is superior to basic information for improving SP authenticity. Since SPs who received an in-depth role-script were perceived as significantly more authentic than SPs with only little information on the patient they were to portray, this method should be established as a standard in SP training.

## **4.3 Indistinguishability of SPs from real patients (*Paper III*)**

Ay-Bryson, D. S., Weck, F., Heinze, P. E., Lang, T., Kühne, F. (in press). Can psychotherapy trainees distinguish standardized patients from real patients? Results of a pilot study. *Zeitschrift für Klinische Psychologie und Psychotherapie*.

### *Theoretical background*

Few studies have examined authenticity of SPs by comparing them with real patients (Krahn et al., 2002; Wünderich et al., 2012). In the often-cited study by Krahn et al. (2002), SPs underwent one training session in which they were provided with a case outline format, which should enable SPs to improvise during simulations, according to the authors. Results of this study showed that most students could correctly distinguish SPs from actual psychiatric patients. In contrast, a more recent study (Wünderich et al., 2012) demonstrated that SPs, following a training of about four hours, could in 60.40% of cases not be detected by experienced psychiatrists. Although it seems plausible that training of SPs may increase authenticity (Partschefeld, 2013), literature provides little evidence for

this assumption. As a consequence, there is no standard on how to train SPs properly. Hence, this study aimed to explore the question as to whether psychotherapy trainees can distinguish between SPs and real patients following an extensive training.

### *Methods*

Psychotherapy trainees participated in this study as raters ( $N = 28$ ;  $M_{\text{age}} = 28.54$  years,  $SD_{\text{age}} = 3.19$ ; 82.14% female) by watching six video-recorded simulations of five-minute therapy segments each. Trainees were asked to assess the interviewee regarding their authenticity using the APD scale and to judge the person's psychological impairment. Finally, since an SP can be rated as authentic, but still not be perceived as a real patient, raters were asked to estimate the likelihood of the interviewee being a real patient. Raters were not informed about the conditions of the interviewee, namely that two of six patients were real patients, while the others were SPs. To enable comparability, all segments were conducted by one licensed psychotherapist and recorded in the laboratory designed as a therapy room. Two real patients were primarily diagnosed with depression (recurrent and current episode), and they were currently undergoing treatment at the outpatient unit of the Department of Clinical Psychology and Psychotherapy at the University of Potsdam. Accordingly, the four SPs were instructed to portray a patient with a first depressive episode. Beforehand, the SPs received a two-day workshop to enable them to simulate their roles authentically (i.e. indistinguishably from real patients), which altogether, including homework, lasted 12 hours.

### *Major findings*

Results of the present study showed that APD mean scores of SPs did not differ significantly from mean scores of real patients  $t(48.93) = -.79$ ,  $p = .43$ , Cohen's  $d = -.21$ , 95% CI [-.75, .33]. Further, raters' estimates of the interviewee to be a real patient did not differ significantly between SPs and real patients  $t(43.79) = .66$ ,  $p = .52$ , Cohen's  $d = .18$ , 95% CI [-.36, .71]. Similarly, ratings of impairment of the interviewed person did not differ between SPs and real patients  $t(51.92) = -.98$ ,  $p = .33$ , Cohen's  $d = -.26$ , 95% CI [-.80, .28]. Finally, a strong correlation was found between authenticity as measured with the APD scale and the likelihood of the interviewed person to be a real patient ( $r = .62 - .76$ ,  $p < .001$ ).

*Conclusion*

In sum, the results of the pilot study imply that psychotherapy trainees could not distinguish between SPs and real patients, and that the more authentic an interviewed person was judged to be, the more likely it was that this person would be predicted to be a real patient. Therefore, it can be concluded that the two-day SP training of the current study was effective considering the authenticity of SPs and hence that thorough SP training is crucial prior to actual simulation. Due to the nature of the pilot study design, further study is needed to replicate these findings based on bigger sample sizes. Limitations and future research directions are discussed in detail in the paper.

## 5. Discussion

The motivation of the present dissertation is based on the observation that the current state of psychotherapy training faces notable issues. First, as outlined above, psychotherapy training as is currently practised cannot be considered evidence-based (Fairburn & Cooper, 2011; Rakovshik & McManus, 2010). Second, trainee perspectives are often neglected and methods to provide trainees with less anxious and more confident learning environments (Pheister et al., 2016) would be desirable. Third, considering the training specifically in Germany, there is demand for more practical methods to ensure a better transition from theory into practice (Strauß & Brähler, 2009).

Although the external validity would be higher with real patient encounters (Kühne et al., 2020), SPs possess a number of advantages over real patients, such as repeatability and standardisation, as I presented previously. Thus, it is not surprising that SPs are currently disseminating more into the field of psychotherapy training (Kühne et al., 2018). At the same time, however, there has been voiced criticism regarding the authenticity of SP portrayals (Kühne et al., 2018). Several authors have argued that the authenticity of SPs is crucial and even more important than consistency (Wind et al., 2004).

Therefore, with the present studies I aimed to examine SP authenticity in order to promote SPs in evidence-based training. In the following, I will summarise the overall contributions the current dissertation has made to research and training based on the findings from *Paper I* to *Paper III* with regard to the previously outlined aims. Subsequently, I will discuss the dissertation's overall limitations and propose future research directions to which the dissertation has led. Finally, I will provide a general conclusion.

### 5.1 Summary of dissertation contributions

Although SPs are increasing in popularity in this field, the authenticity of their portrayals has not been studied thoroughly. Studies that did previously explore authenticity of SPs portraying mental health problems, such as obsessive-compulsive disorder, substance abuse or depression (Fussell et al., 2009; Partschefeld, 2013; Partschefeld et al., 2013; Wüdrich et al., 2008) were limited to small sample sizes, lacked randomisation, control groups, pre- and post-measures, or were limited to female-only SPs. Further, authenticity

was mainly measured using single-items or open questions instead of a previously validated tool. Hence, based on a total of three empirical studies, I pursued three main aims to uncover how authenticity can be assessed by developing and validating an observer-based rating instrument and, subsequently, how authenticity can be enhanced.

*Aim 1: Develop and validate an observer-based rating scale that can be used to assess SPs' authenticity, specifically designed for psychotherapeutic settings.*

Since, to our knowledge, no instrument was available with which the authenticity of SPs in psychotherapeutic contexts can be measured, in the first study the APD scale was developed and validated. Based upon preliminary evidence of the paper (Ay-Bryson et al., 2020), the APD scale has been demonstrated to be a valid and reliable tool. For this reason, the APD scale was deployed in the follow-up studies of the dissertation.

When I first developed the research proposal to study SP authenticity, I came across two potentially relevant instruments that pre-existed: the Maastricht assessment of Simulated Patients (MaSP; Wind et al., 2004) and the Nijmegen Evaluation of the Simulated Patient (NESP; Bouter et al., 2013), both of which stem from medical education. The MaSP was constructed with the aim “to improve the evaluation of SP performance” (p. 40) and consists of two subscales to measure (1) the authenticity during the consultation and (2) the feedback after the consultation. The first subscale refers to the SP’s portrayal during the simulation, whereas the second refers to the quality of the SP’s comments provided to the student-doctor after the interaction. While the authors of the MaSP found an acceptable internal consistency, they also found that twelve completed evaluations were needed to obtain a reliable evaluation for one SP. Years later, Bouter and colleagues (2013) constructed the NESP. In addition to the subscales “role-play” and “process of feedback”, the authors added a third subscale “application of feedback rules”. Bouter et al. (2013) justified their purpose by pointing to the necessity of assessing “the quality of the SP’s role-playing and feedback abilities within the context of giving feedback from a patient perspective while also focussing on students’ communication skills and medical knowledge” (p. 258). Both of these instruments, i.e. MaSP and NESP, were initially considered for measuring authenticity in the present studies. However, both measures did not meet the required criteria that were needed for the purpose of my dissertation. The first aim of this project was not to construct a valid and reliable tool to

assess the authenticity when performing *any* patient. Since I was interested in SPs for psychotherapy training, SPs portraying primarily medical patients were not part of my scope. A tool designed to measure authenticity of a psychotherapy patient had to be tailored specifically for the psychotherapeutic context. Further, we wanted to refrain from the term “SP” in the items as we did not want to reveal the SP status per se in order to operationalise the third research question of the dissertation. Therefore, while still considering both measures during our item construction, we developed the APD scale. Importantly, the APD scale was not designed disorder-specific. Consequently, the APD scale can be used to assess the portrayal of various mental disorders, though it should be noted that the instrument has thus far only been validated on portrayals of depression.

All in all, the first aim was fulfilled as there now exists an observer-based rating instrument that can be used to assess SP authenticity (a) in psychotherapy contexts, and (b) that can, theoretically, be used for any mental disorder. With the APD scale we hope to foster further dissemination of SPs into evidence-based training of psychotherapy.

*Aim 2: Study whether student SPs can portray a mental health problem more authentically after receiving an in-depth role-script compared to a baseline instruction.*

The second paper addresses the second aim of the dissertation. The results of this study demonstrated that student SPs who received more information on a patient case (i.e. in-depth role-script), were perceived to be significantly more authentic compared to those SPs who only received basic information. Accordingly, when using SPs for psychotherapy research and training, SPs should receive proper patient descriptions beforehand.

These results are in line with previous studies (Fussell et al., 2009; Partschefeld, 2013; Wünderich et al., 2008). All of these studies reported that their SPs were perceived authentic in portraying a mental disorder. Hence, with the current study results we too can confirm the evidence that SPs can be trained in order to authentically portray a patient with a mental disorder.

In the study by Wünderich et al. (2008), eight SPs were recruited via personal contacts. Three of the SPs were caregivers with experience in the subject of psychiatry, four were psychology interns, and one SP was an amateur actor. All SPs received individual training; the procedure of this was not further specified by the authors. The role-scripts

in this study contained information regarding age, external features, characteristics as well as a description of the current complaint, psychiatric history, and biography. In order to facilitate authenticity, the authors reportedly incorporated literal verbalisations. The SPs received 12.50€ per hour. Finally, medical students evaluated the SPs. Although the SPs were assessed in the subject of psychiatry, the results are only partly comparable, as the evaluator perspectives between this (Wünderich et al., 2008) and our study differ (medical vs. psychological perspectives). Nevertheless, our results are in line with this study (Wünderich et al., 2008) in that the SPs were found to be surprisingly authentic.

In the pilot study of Fussell et al. (2009), two professional actors were examined. The SPs of this study were trained for 15 to 25 hours. The role-script the authors of this study used included: social, medical, family, and drug history, as well as symptoms, a timeline of co-morbidities, work, insurance and relationships with others. The authenticity of the SPs was evaluated by practising clinicians as well as substance abuse counselling students. Although the authors did not specify how much the SPs were paid, it is likely that reimbursement of professional actors is higher compared to students. Thus, the current study (Ay-Bryson et al., minor revision) expands our insights in this field by demonstrating that student SPs are a promising alternative to professional actors, as they were perceived authentic based on the portrayal of depression. Compared to the study of Fussell et al. (2009), in which SPs portrayed patients with substance abuse, at this time, we cannot confirm that students SPs are suitable for this kind of disorder as well, and thus need further study on this.

Finally, in the study of Partschefeld (2013), the SPs (four Psychology students and four teenagers) were involved in the development of role-scripts. In her study, Partschefeld argued that SPs were capable of more authentically portraying a patient because they were involved in script planning. The role-scripts of the present study (Ay-Bryson et al., minor revision), nevertheless, were previously developed by a licensed therapist and SPs received them as the experiment took place. Hence, this new insight suggests, in contrast to Partschefeld (2013), that SPs do not necessarily have to be involved in script-planning. This may promote a more cost-effective approach, a quality in high demand (Lane & Rollnick, 2007), to the SP methodology, as SPs would not be required to be engaged prior to the simulation per se. Nevertheless, based on this study I am not able to draw any conclusion on whether SPs would be even more authentic if they were involved in script planning, which may be worth exploring in future studies.



All in all, considering the vast heterogeneity of implemented role-scripts or instructions for SPs, this – to the best of my knowledge – is the first study that, systemically and within a randomised controlled study design, demonstrated that (student) SPs profit from detailed information on a patient case compared to basic information. Further, the observations of this study (Ay-Bryson et al., minor revision) demonstrated that small details, such as nonverbal behaviour, contribute to a more authentic portrayal. For instance, in the paper we discuss that the raters of the study remarked that SPs talked more slowly and quietly after receiving more details, and that they held a different posture. This observation is in line with the conclusion of Wünderlich et al. (2012), who highlighted that “intricacies in the simulation (may) potentially reduce authenticity (...)” (p. 501). The authors noted that details, such as “gestures too quick for depression” or “did not present enough formal thought disorders” (p. 501), hindered an authentic portrayal. Accordingly, when planning role-scripts for SPs, particular attention should be paid towards details.

*Aim 3: Investigate whether SPs can be distinguished from real patients by psychotherapy trainees following an exhaustive SP training.*

The third paper addresses the third aim. Since the dissertation’s superior aim is to offer a beneficial contribution to psychotherapy training and research, the perspective of psychotherapy trainees is of particular interest. Therefore, the third aim of this research was to investigate whether trainees are capable of distinguishing real patients from SPs who have undergone an extensive two-day SP training. The results of this study revealed that psychotherapy trainees were not able to distinguish between real patients and SPs. In fact, descriptive results even showed that the most authentic person happened to be an SP.

The present results contradict the results of Krahn et al. (2002). In the study of Krahn et al. (2002), standardised outpatients received a psychiatric case that was based on a real patient and written by a psychiatrist. Medical students evaluated the SPs as less authentic compared to real patients. In addition, the authors reported that the students felt less empathy towards the SPs. In the current study (Ay-Bryson et al., in press) we cannot deliver results regarding empathy. However, we do demonstrate that the trainees of our study did not find real patients more authentic per se. The results, however, are only

partially comparable: in the study of Krahn et al. (2002), medical students evaluated SPs, while in our study psychotherapy trainees assessed SPs. Krahn et al. described that their SP training included handouts, videotaped cases, and observed performances. The duration of the training, however, was not mentioned. The authors concluded that “training must focus on facilitating actors’ ability to convey emotion realistically and therefore evoke empathy in the interviewer” (p. 30). Although we cannot narrow down which of the training components of the present study (Ay-Bryson et al., in press) was most helpful for the SPs, the SP training included authenticity, psychoeducation, role description and role analysis. Hence, it can be assumed that SPs were trained in the emotional part of the depressive case they were to portray, and can thus, with caution, confirm Krahn et al.’s (2002) conclusion.

Finally, the current results (partly) disagree with another study that compared the authenticity of SPs and real patients (Wüdrich et al., 2012). In this study, the authors concluded that SPs who received four hours of training to develop authentic acting skills were not as authentic as real patients. Nevertheless, they also underlined that in most cases SPs were not detected, which would be in line with our results. However, since the authors defined authenticity as the impossibility to distinguish between SPs and real patients (Wüdrich et al., 2012), it comes as a surprise that the authors concluded that their SPs were not authentic, even though experienced psychiatrists were mostly unable to distinguish between SPs and real patients. Regardless of this, both studies (Wüdrich et al., 2012; Ay-Bryson et al., in press) demonstrated that SPs could not be detected. This result indicates that experienced clinicians evaluate authenticity in a similar way to novice psychotherapy trainees. However, to draw firm conclusions, further study on this is needed.

In short, this study (Ay-Bryson et al., in press) demonstrated that psychotherapy trainees are not able to tell real patients from SPs who had undergone an extensive SP training including thorough discussion on the clinical picture of depression.

In summary, the dissertation contributes the following main new and supporting insights to the literature on SPs:

- (1) Authenticity of SPs employed in psychotherapy training can be assessed using the newly-developed and validated APD scale (*Paper I*);

- (2) Thorough planning of role-scripts appears to pay off as student SPs evidently profit from detailed role-scripts regarding the authentic portrayal of a depressive patient compared to basic information (*Paper II*);
- (3) Students seem to be a promising demographic from which to recruit SPs (*Paper II*);
- (4) Role-scripts rich in detail, including details regarding nonverbal behaviour, should be included in SP training when possible (*Paper II*);
- (5) Whenever possible, SPs should receive an extensive training including a broad introduction to the clinical picture, authenticity as well as role-play exercises (*Paper III*);
- (6) Following proper recruitment and training, SPs in general cannot be distinguished from real patients by psychotherapy trainees (*Paper III*).

## **5.2 Limitations and future research directions**

Although the dissertation has made some contributions to research with SPs, which will hopefully encourage the improvement of psychotherapy training, this research has some limitations that are also important to be discussed. In this context, the dissertation leads to a number of further questions to be explored. Based on these, future research directions will be outlined in the following. Note, however, that study-specific limitations are discussed in the papers directly and will therefore not be repeated wholly here.

### **5.2.1 Student SPs**

In the present studies, only student SPs were considered. Nonetheless, as was mentioned previously, there is a variety of terms referring to the concept of SPs (Kühne et al., 2018; Nestel & Bearman, 2015). Accordingly, various people can engage as an SP, such as laypersons, faculties, actors, or indeed students. While student SPs offer a cost-effective approach, there may be circumstances under which student SPs are not appropriate. For instance, student SPs may be acquainted with the trainees, which would then diminish the superiority of the SP methodology over role-plays with fellow students. Further, the number of students who want to engage as SPs may be limited, for which reason a broader database of SPs should be aimed for. Finally, and arguably most importantly, only student SPs would not represent a realistic image of real clinical encounters. Clinicians encounter patients of all ages, professions and backgrounds and are not to be narrowed down to

students. Thus, for these clinical as well as logistical reasons, the present study results should be expanded to SPs with various backgrounds.

### **5.2.2 Rater perspective and training**

Similarly, in the present studies we mainly engaged students or trainees as evaluators of SPs. In *Paper I*, the majority of evaluators were psychotherapy trainees; in *Paper II*, we administered four perspectives for the evaluation of SPs (i.e. students, study therapist, licensed therapist and SPs), and in *Paper III* we again asked psychotherapy trainees to evaluate authenticity. With regard to the evaluator perspective, in the literature on therapeutic competence, the following questions are often posed (Weck, 2013): Who is eligible to accurately evaluate therapist competence? How much information should be included for a stable assessment? In his book, Weck (2013) describes that the evaluation (of therapist competence) made by independent observers displays a gold standard as it provides the most objective appraisal. At the same time, Weck (2013) and Muse and McManus (2013) notice that there has been a lack of clear guidelines regarding raters' qualifications and the rater training. Literature regarding more complex therapeutic behaviours suggests that evaluators with more clinical experience should be able to adequately assess therapist competence unlike inexperienced raters (Weck, 2013). Similar to the assessment of therapist competence, it is conceivable that more experienced clinicians will more reliably evaluate SP authenticity. On the other hand, it may be argued that the trainee perspective is more important to consider as they are being trained with SPs.

Even though the evaluator perspective of SP authenticity was not the subject of my dissertation, based on the three papers the dissertation provides preliminary results with regard to the evaluator perspective. In *Paper II* (Ay-Bryson et al., minor revision), the agreement between the student raters and the study therapist was good, while the agreement between the student raters and the licensed therapist as well as between the study therapist and the licensed therapist was only moderate in comparison. This result may indicate that the evaluations between trainees and more experienced clinicians may diverge with regard to SP authenticity. On the other hand, the results of *Paper III* (Ay-Bryson, et al., in press) demonstrated that psychotherapy trainees could not distinguish between SPs and real patients, which is in line with a former study (Wünderlich et al., 2012) in which experienced psychiatrists could in most cases not detect an SP. Both of these

studies may be indicating that clinical experience is of lesser importance when assessing authenticity. Nevertheless, since little is known on this, future studies are needed to consider various evaluator perspectives. Hence, in addition to the suggestion made by Wüdrich et al. (2012) that future studies should investigate “what exact degree of authenticity is needed to make teaching with SPs effective” (p. 502), I propose that the degree of authenticity should be studied in relation with the evaluator perspective.

Moreover, the APD scale is not designed disorder-specific. As a consequence, whoever applies the APD needs a certain degree of knowledge on or expectation of “how a typical patient” looks, talks or behaves. In the current studies, for instance, the raters had to have a picture of a “typical depressive patient” as the SPs were portraying depression. Even though the raters of *Paper II* (Ay-Bryson et al., minor revision) were trained in accordance with literature guidelines (Feldman et al., 2012), the other two studies did not include a prior rater training. One implication may be that a prior training is not always necessary. Nevertheless, it should be systematically examined whether and to what extent rater training is required to evaluate authenticity and to achieve high degrees of agreement between independent observers.

### **5.2.3 *Portrayal of depression only***

The current dissertation is limited to the portrayal of depression only. However, since depression is among the most commonly diagnosed mental disorders for adults (Jacobi et al., 2014; Richards, 2011), it is highly likely for trainees to encounter depressive patients during their training. Therefore, depression may be considered crucial to examine in the context of SP authenticity. It is nonetheless essential to replicate the results of the current dissertation based on the portrayal of other mental disorders as well. Importantly, the portrayal of more complex disorders should be in the scope of future studies. Complex mental disorders, such as mania or psychosis, are considered to be more difficult for SPs to portray (Kühne et al., 2018). Simultaneously, trainees are likely to need training particularly with more complex (and comorbid) patients as these encounters may be perceived as more challenging.

### **5.2.4 *Lack of SP distress monitoring***

Another limitation of the dissertation that should be mentioned is that we did not control for potential distress of the SPs in the present studies. Hence, we cannot deliver insight

on a potential relation between authenticity and distress experienced by the SPs. For preventative measures, it would be desirable to study whether a more authentic portrayal causes more distress in SPs. For instance, one study reported that the SPs occasionally reported back that they found their role to be heavy and that in some cases they experienced symptoms due to their SP performances (Bokken et al., 2004). With this in mind, in the second study (Ay-Bryson et al., minor revision) we provided a de-roling technique (Kühne et al., 2018) as well as debriefing, during which we ensured the participant was well before ending the experiment. Nonetheless, the potential risk of stress occurrences due to SP performance as in the study of Bokken et al. (2004) should be taken seriously. According to the present dissertation, the more information the SPs receive on the patient role, the more authentic their simulation is perceived. It might therefore be conceivable that SPs become more distressed the more time they spend with their role to be simulated, and the more information they receive. Therefore, future research should examine whether more authenticity (i.e. more involvement with the character) leads to more perceived stress or a decreased well-being in the SPs.

### ***5.2.5 Lack of relation between authenticity and learning effectiveness***

As was outlined previously, many authors have highlighted the importance of authenticity in the SP methodology (Kühne et al., 2018). In some cases, authenticity was even valued as more important than consistency of performances (Wind et al., 2004). Some authors have indicated that an authentic portrayal is required to make learning with SPs effective. Otherwise, the learning effect may be dampened as the study of Krahn et al. (2001) demonstrated. One systematic review (Issenberg et al., 2005) directly examined the relation between authenticity or 'high fidelity' (as the authors call it) and learning effects. The authors found that high-fidelity simulations in medical education were educationally effective. Their results supported the assumption that high fidelity results in enhanced learning. Although this implication may also appear plausible, other authors disagreed with this (Hamstra et al., 2014; Norman et al., 2012). In their paper, Norman et al. (2012) compared high- with low-fidelity simulations in relation with clinical performances and they found that both led to comparable learning effects. In fact, they did not find higher fidelity to be superior. Therefore, it does not come as a surprise that the authors did not agree with the opinion that high fidelity is of paramount importance, especially as they argue that high-fidelity simulation results in higher costs. Importantly, however, it needs

to be noted that the criticism stems from medical education, where simulation often contains mannequins, which may contribute to the high costs referred to by Norman et al. (2012). At the same time, authenticity may play more of an important role in the context of psychotherapy training, as opposed to medical education. It is conceivable that simulating a psychotherapy patient with a formal thought disorder is more challenging than, for instance, simulating a patient with back pain seeking physiotherapy. Further, whether higher SP authenticity would result in higher costs in the context of psychotherapy training remains debatable at this point in time. Nevertheless, since the present dissertation did not examine the relation between authenticity and learning effect, we cannot draw any conclusion on this possibility. Therefore, it would be desirable for future studies to study the relation in question (Hamstra et al., 2014; Norman et al., 2012) between learning effects and authenticity, adapted to psychotherapy training.

### 5.3 General discussion

Based on a total of three empirical studies, the current dissertation has demonstrated that SPs can be trained in order to authentically portray a patient with a mental disorder, i.e. depression. Smith et al. (2015) acknowledged that there is not only one way to train SPs for two reasons: first, training may vary greatly depending on the pre-experience of the SPs; second, the training may differ depending on the profession and the ultimate goal. From the present studies, however, we can conclude that high levels of authenticity for the portrayal of depression (i.e. indistinguishability from real patients) can be achieved with an exhaustive two-day training as described in *Paper III*. For institutions with fewer resources where there is a need for cheaper alternatives (Lane & Rollnick, 2007), student SPs can firstly be engaged, and secondly, they may be trained with detailed in-depth role-scripts, which evidently increase authenticity (*Paper II*).

The results further show that attention should be paid to details, such as the emotional reaction and nonverbal behaviour of an SP (Ay-Bryson et al., minor revision; Wüdrich et al., 2012); overly specific responses should be avoided as they might appear rehearsed and might not represent realistic patient reactions (Fussell et al., 2009). Similarly, Dieckmann and colleagues (2007) conducted interviews with anaesthesiologists regarding their perception of realism in patient simulations they had encountered. The authors found plausibility to play a core element in the perception of realism. This means that when participants felt that the simulation could have happened in real life, i.e. the

simulation was plausible, they felt that it was realistic, while an exaggerated simulation led to a fictive perception of the simulation.

#### **5.4 SPs and evidence-based training**

Ultimately, the superior goal of psychotherapy research and training research is to provide better care for patients suffering from a mental health problem. Therefore, it is essential to shift psychotherapy training more into the focus of research as it depicts a basis for the standard care of patients (Weck, 2017). In the beginning of the dissertation I argued that it might be possible that psychotherapists need effective education in order to deliver treatment properly. In a similar way it can be argued that SP training needs to be evidence-based in order to offer effective teaching methods. While there is still much room for improvement and some questions remain unanswered, the present dissertation reveals promising results with practical implications that will hopefully promote evidence-based SP training. Considering the many positive effects of SPs on learning and their various other advantages, this approach represents an encouraging and promising way not only to train psychotherapy trainees effectively, but also to address trainees' perceived anxieties and challenges during first patient encounters (Hill et al., 2007).

Looking at Germany, where the recently reformed psychotherapy law prescribes the use of SPs in the curriculum, it is conceivable that SPs programmes will become more prevalent in the upcoming years. First papers on the integration of SPs in universities are already available and they report an overall favourable feasibility (Kühne et al., in press) and increased self-efficacy in students trained with SPs (Alpers & Steiger-White, 2020). It will be highly interesting to examine whether this practice-oriented teaching method will eventually result in better patient outcomes, too.

#### **5.5 Conclusion**

Altogether, the present dissertation has contributed to research on SPs in the context of psychotherapy training. The results of the herein reported studies have several implications for future research on the authenticity of SPs as well as training programmes that anticipate integrating SPs into their curriculum. It is hoped that this dissertation will be found useful for future studies and will encourage examinations of SP authenticity in the context of different mental health problems.



In conclusion, the present work suggests that objective and reliable assessments of SP authenticity can be obtained with the APD scale. Further, extensive training for SPs should be considered, whenever possible. Through training SPs for an authentic portrayal, we may contribute to the improvement of evidence-based training, which will hopefully, in turn, improve psychotherapy. Lastly, careful planning and writing of role-scripts rich in detail seems to be vital for authenticity.

Intriguingly, an award-winning actor comes to the same conclusion:

“Everything at the end of the day (when you’re making a movie) starts with the material and at how well written the script is.” – Leonardo DiCaprio

## 6. References

- Abdool, P. S., Nirula, L., Bonato, S., Rajji, T. K., & Silver, I. L. (2017). Simulation in undergraduate psychiatry: Exploring the depth of learner engagement. *Academic Psychiatry, 41*(2), 251–261. <https://doi.org/10.1007/s40596-016-0633-9>
- Adamo, G. (2003). Simulated and standardized patients in OSCEs: Achievements and challenges 1992-2003. *Medical Teacher, 25*(3), 262–270. <https://doi.org/10.1080/0142159031000100300>
- Alpers, G. W., & Steiger-White, F. (2020). Simulationspatienten im universitären Psychologiestudium: Einführung einer innovativen Methode zur Stärkung praxisorientierter Lehre in klinischer Psychologie und Psychotherapie. *Verhaltenstherapie, 30*, 104–116. <https://doi.org/10.1159/000507740>
- Ay-Bryson, D. S., Weck, F., Heinze, P. E., Lang, T., & Kühne, F. (in press). Can psychotherapy trainees distinguish standardized patients from real patients? Results of a pilot study. *Zeitschrift für Klinische Psychologie und Psychotherapie*.
- Ay-Bryson, D. S., Weck, F., & Kühne, F. (2020). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000349>
- Ay-Bryson, D. S., Weck, F., & Kühne, F. (minor revision). Can students in simulation portray a psychotherapy patient authentically with a detailed role-script? Results of a randomized-controlled study. *Training and Education in Professional Psychology*.
- Baldwin, S. A., & Imel, Z. E. (2013). Therapist Effects: Findings and Methods. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (Vol. 6, pp. 258–297). John Wiley & Sons.
- Barkham, M., Lutz, W., Lambert, M. J., & Saxon, D. (2017). Therapist Effects, Effective Therapists, and the Law of Variability. In L. G. Castonguay and C. E. Hill (Eds.), *How and why are some therapists better than others? Understanding therapist effects* (pp. 13–36). American Psychological Association. <http://dx.doi.org/10.1037/0000034-000>

- Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Academic Medicine*, 68(6), 443–451. <https://doi.org/10.1097/00001888-199306000-00002>
- Barrows, H. S., & Abrahamson, S. (1964). The programmed patient: A technique for appraising student performance in clinical neurology. *Journal for Medical Education*, 39(8), 802–805.
- Bearman, M., & Nestel, D. (2015). Learning theories and simulated patient methodology. In D. Nestel and M. Bearman (Eds.), *Simulated Patient Methodology: Theory, Evidence and Practice* (pp. 33–38). John Wiley & Sons.
- Bokken, L., Van Dalen, J., & Rethans, J.-J. (2004). Performance-related stress symptoms in simulated patients. *Medical Education*, 38(10), 1089–1094. <https://doi.org/10.1111/j.1365-2929.2004.01958.x>
- Bouter, S., van Weel-Baumgarten, E., & Bolhuis, S. (2013). Construction and validation of the nijmegen evaluation of the simulated patient (nesp): Assessing simulated patients' ability to role-play and provide feedback to students. *Academic Medicine*, 88(2), 253–259. <https://doi.org/10.1097/ACM.0b013e31827c0856>
- Callahan, J. L., & Watkins, C. E., Jr. (2018). Evidence-based training: The time has come. *Training and Education in Professional Psychology*, 12(4), 211–218. <https://doi.org/10.1037/tep0000204>
- Castonguay, L. G., Boswell, J. F., Constantino, M. J., Goldfried, M. R., & Hill, C. E. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist*, 65(1), 34–49. <https://doi.org/10.1037/a0017330>
- Charité Universitätsmedizin Berlin. (2020). *Charité Simulationspatientenprogramm*. [https://www.charite.de/studium\\_lehre/struktur\\_einrichtungen/simulationspatientenprogramm/](https://www.charite.de/studium_lehre/struktur_einrichtungen/simulationspatientenprogramm/)
- Cook, D. A., Brydges, R., Hamstra, S. J., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hatala, R. (2012). Comparative effectiveness of technology-enhanced simulation versus other instructional methods: A systematic review and meta-analysis. *Simulation in Healthcare*, 7(5), 308–320. <https://doi.org/10.1097/SIH.0b013e3182614f95>
- Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hamstra, S. J. (2011). Technology-enhanced simulation for health

- professions education: A systematic review and meta-analysis. *Jama*, 306(9), 978–988. <https://doi.org/10.1001/jama.2011.1234>
- Coyle, B., Miller, M., & McGowen, K. R. (1998). Using standardized patients to teach and learn psychotherapy. *Academic Medicine*, 73(5), 591–592. <https://doi.org/10.1097/00001888-199805000-00058>
- Davies, E. L., Pelentsov, L. J., Montagu, A., Gordon, A. L., Hooper, K. J., & Esterman, A. J. (2020). ‘Who Am I and Why Am I Here?’ A Scoping Review Exploring the Templates and Protocols That Direct Actors in Their Roles as Simulated (Standardized) Patients. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*.
- Dennhag, I., Gibbons, M. B. C., Barber, J. P., Gallop, R., & Crits-Christoph, P. (2012). How many treatment sessions and patients are needed to create a stable score of adherence and competence in the treatment of cocaine dependence? *Psychotherapy Research*, 22(4), 475–488. <https://doi.org/10.1080/10503307.2012.674790>
- Entwurf eines Gesetzes zur Reform der Psychotherapeutenausbildung, no. 19/9770 (2019). [https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/Gesetze\\_und\\_Verordnungen/GuV/P/Psychotherapeutenausbildung\\_Reform\\_Bundestag.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/P/Psychotherapeutenausbildung_Reform_Bundestag.pdf)
- Dieckmann, P., Manser, T., Wehner, T., & Rall, M. (2007). Reality and fiction cues in medical patient simulation: An interview study with anesthesiologists. *Journal of Cognitive Engineering and Decision Making*, 1(2), 148–168. <https://doi.org/10.1518/155534307X232820>
- Eagles, J. M., Calder, S. A., Wilson, S., Murdoch, J. M., & Sclare, P. D. (2007). Simulated patients in undergraduate education in psychiatry. *Psychiatric Bulletin*, 31(5), 187–190. <https://doi.org/10.1192/pb.bp.106.010793>
- Eckel, J., Merod, R., Vogel, H., & Neuderth, S. (2014). Einsatz von Schauspielpatienten in den „Psych-“Fächern des Medizinstudiums–Verwendungsmöglichkeiten in der Psychotherapieausbildung? *PPmP-Psychotherapie· Psychosomatik· Medizinische Psychologie*, 64(01), 5–11. <https://doi.org/10.1055/s-0033-1347210>
- Eid, A., Petty, M., Hutchins, L., & Thompson, R. (2009). “Breaking bad news”: Standardized patient intervention improves communication skills for hematology-

- oncology fellows and advanced practice nurses. *Journal of Cancer Education*, 24(2), 154–159. <https://doi.org/10.1080/08858190902854848>
- Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy*, 49(6–7), 373–378. <https://doi.org/10.1016/j.brat.2011.03.005>
- Feldman, M., Lazzara, E. H., Vanderbilt, A. A., & DiazGranados, D. (2012). Rater training to support high-stakes simulation-based assessments. *Journal of Continuing Education in the Health Professions*, 32(4), 279–286. <https://doi.org/10.1002/chp.21156>
- Fussell, H. E., Shannon-Lewy, C., & McFarland, B. (2009). Evaluating and training substance abuse counselors: A pilot study assessing standardized patients as authentic clients. *Substance Abuse*, 30(1), 47–60. <https://doi.org/10.1080/08897070802606360>
- Hagell, P., Hedin, P.-J., Meads, D. M., Nyberg, L., & McKenna, S. P. (2010). Effects of method of translation of patient-reported health outcome questionnaires: A randomized study of the translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden. *Value in Health*, 13(4), 424–430. <https://doi.org/10.1111/j.1524-4733.2009.00677.x>
- Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine*, 89(3), 387–392. <https://doi.org/10.1097/ACM.0000000000000130>
- Harden, R. M., Stevenson, M., Downie, W. W., & Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, 1(5955), 447–451. <https://doi.org/10.1136/bmj.1.5955.447>
- Hill, C. E., & Knox, S. (2013). Training and Supervision in Psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (Vol. 6, pp. 775–811). John Wiley & Sons.
- Hill, C. E., & Lent, R. W. (2006). A narrative and meta-analytic review of helping skills training: Time to revive a dormant area of inquiry. *Psychotherapy: Theory, Research, Practice, Training*, 43(2), 154–172. <https://doi.org/10.1037/0033-3204.43.2.154>

- Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivlighan Jr, D. M., & Gelso, C. J. (2017). Therapist expertise in psychotherapy revisited. *The Counseling Psychologist*, 45(1), 7–53. <https://doi.org/10.1177/0011000016641192>
- Hill, C. E., Sullivan, C., Knox, S., & Schlosser, L. Z. (2007). Becoming psychotherapists: Experiences of novice trainees in a beginning graduate class. *Psychotherapy: Theory, Research, Practice, Training*, 44(4), 434–449. <https://doi.org/10.1037/0033-3204.44.4.434>
- Hoffmann, S. O., Rudolf, G., & Strauß, B. (2008). Unerwünschte und schädliche Wirkungen von Psychotherapie. *Psychotherapeut*, 53, 4–16. <https://doi.org/10.1007/s00278-007-0578-2>
- Imel, Z. E., Baldwin, S. A., Baer, J. S., Hartzler, B., Dunn, C., Rosengren, D. B., & Atkins, D. C. (2014). Evaluating therapist adherence in motivational interviewing by comparing performance with standardized and real patients. *Journal of Consulting and Clinical Psychology*, 82(3), 472–481. <https://doi.org/10.1037/a0036158>
- Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher*, 27(1), 10–28. <https://doi.org/doi:10.1080/01421590500046924>
- Jacobi, F., Höfler, M., Strehle, J., Mack, S., Gerschler, A., Scholl, L., Busch, M. A., Maske, U., Hapke, U., Gaebel, W., Maler, W., Wagner, M., Zielasek, J., & Wittchen, H.-U. (2014). Psychische Störungen in der Allgemeinbevölkerung. *Der Nervenarzt*, 85, 77–87. <https://doi.org/10.1007/s00115-013-3961-y>
- Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review*, 67, 78–93. <https://doi.org/10.1016/j.cpr.2018.08.004>
- Kevala, A., & Moosbrugger, H. (2012). *Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilung: Vol. 2., aktualisierte und überarbeitete Auflage*. Springer-Verlag.
- Klamen, D. L., & Yudkowsky, R. (2002). Using standardized patients for formative feedback in an introduction to psychotherapy course. *Academic Psychiatry*, 26(3), 168–172. <https://doi.org/doi:10.1176/appi.ap.26.3.168>

- Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2001). Experiential learning theory: Previous research and new directions. In R. J. Sternberg and L.-F. Zhang (Eds.), *Perspectives on thinking, learning, and cognitive styles* (Vol. 1, pp. 227–247). Lawrence Erlbaum Associates Publishers.
- Krahn, L. E., Bostwick, J. M., Sutor, B., & Olsen, M. W. (2002). The challenge of empathy. *Academic Psychiatry*, *26*, 26–30. <https://doi.org/10.1176/appi.ap.26.1.26>
- Kühne, F., Maaß, U., & Weck, F. (in press). Einsatz standardisierter Patienten im Psychologiestudium: Von der Forschung in die Praxis. *Verhaltenstherapie*. <https://doi.org/10.1159/000509249>
- Kühne, F., Ay, D. S., Otterbeck, M., & Weck, F. (2018). Standardized patients in clinical psychology and psychotherapy: A scoping review on barriers and facilitators for implementation. *Academic Psychiatry*, *42*(6), 773–781. <https://doi.org/10.1007/s40596-018-0886-6>
- Kühne, F., Heinze, P. E., & Weck, F. (2020). Standardized patients in psychotherapy training and clinical supervision: Study protocol for a randomized controlled trial. *Trials*, *21*(276), 1–7. <https://doi.org/10.1186/s13063-020-4172-z>
- Lambert, M. J. (2013). The Efficacy and Effectiveness of Psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (Vol. 6, pp. 169–218). John Wiley & Sons.
- Lane, C., & Rollnick, S. (2007). The use of simulated patients and role-play in communication skills training: A review of the literature to August 2005. *Patient Education and Counseling*, *67*(1–2), 13–20. <https://doi.org/10.1016/j.pec.2007.02.011>
- Lange, E., Hillebrand, V., & Pfäfflin, F. (2009). Beschwerden über Therapeuten. *Psychotherapeut*, *54*, 307–309. <https://doi.org/10.1007/s00278-009-0682-6>
- Lewis, M., Bell, J., & Asghar, A. (2008). Use of simulated patients in development of physiotherapy students' interpersonal skills. *International Journal of Therapy and Rehabilitation*, *15*(5), 221–229. <https://doi.org/10.12968/ijtr.2008.15.5.29234>
- Lutz, W., & Barkham, M. (2015). Therapist Effects. In R. Cautin & S. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology* (pp. 1–6). Wiley-Blackwell. <https://doi.org/10.1002/9781118625392.wbecp109>

- May, W., Park, J. H., & Lee, J. P. (2009). A ten-year review of the literature on the use of standardized patients in teaching and learning: 1996–2005. *Medical Teacher, 31*(6), 487–492. <https://doi.org/10.1080/01421590802530898>
- McGaghie, W. C., Issenberg, S. B., Cohen, M. E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Academic Medicine: Journal of the Association of American Medical Colleges, 86*(6), 706. <https://doi.org/10.1097/ACM.0b013e318217e119>
- McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*(2), 73–84. <https://doi.org/10.1037/a0018121>
- McMahon, A., & Ledden, K. (2019). Recording client sessions during psychotherapy training: From “an absurd idea” to “potent learning”. *Counselling and Psychotherapy Research, 19*(3), 241–251. <https://doi.org/10.1002/capr.12209>
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), S63-7. <https://doi.org/10.1097/00001888-199009000-00045>
- Murtagh, G. M. (2015). Simulated interaction and authentic interaction—A place for Conversation Analysis? In D. Nestel and M. Bearman (Eds.), *Simulated Patient Methodology: Theory, Evidence and Practice* (pp. 46–52). John Wiley & Sons.
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review, 33*(3), 484–499. <https://doi.org/10.1016/j.cpr.2013.01.010>
- Nerdrum, P., & Rønnestad, M. H. (2002). The trainees’ perspective: A qualitative study of learning empathic communication in Norway. *The Counseling Psychologist, 30*(4), 609–629. <https://doi.org/10.1177/00100002030004007>
- Nestel, D., Fleishman, C., & Bearman, M. (2015). Preparation: Developing scenarios and training for role portrayal. In D. Nestel and M. Bearman (Eds.), *Simulated Patient Methodology: Theory, Evidence and Practice* (pp. 63–70). John Wiley & Sons.
- Nestel, D., & Bearman, M. (2015). Introduction to simulated patient methodology. In D. Nestel and M. Bearman (Eds.), *Simulated Patient Methodology: Theory, Evidence and Practice* (pp. 1–4). John Wiley & Sons.



- Nestel, D., Morrison, T., & Pritchard, S. (2015). Scope of contemporary simulated patient methodology. In Nestel and M. Bearman (Eds.), *Simulated Patient Methodology: Theory, Evidence and Practice* (pp. 7–15). John Wiley & Sons.
- Newble, D. (2004). Techniques for measuring clinical competence: Objective structured clinical examinations. *Medical Education*, 38(2), 199–203. <https://doi.org/10.1111/j.1365-2923.2004.01755.x>.
- Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education*, 46(7), 636–647. <https://doi.org/10.1111/j.1365-2923.2012.04243.x>
- Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology*, 62(9), 1157–1172. <https://doi.org/10.1002/jclp.20272>
- Partschefeld, E. (2013). *Evaluation des Einsatzes von Simulationspatienten in der psychotherapeutischen Ausbildung: Ein Beitrag zur empirischen Fundierung der Psychotherapieausbildung* [Doctoral dissertation].
- Partschefeld, E., Strauß, B., Geyer, M., & Philipp, S. (2013). Simulationspatienten in der Psychotherapieausbildung. *Psychotherapeut*, 58(5), 438–445. <https://doi.org/10.1007/s00278-013-1002-8>
- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3), 423–451. [https://doi.org/10.1207/s15327809jls1303\\_6](https://doi.org/10.1207/s15327809jls1303_6)
- Perera, J., Perera, J., Abdullah, J., & Lee, N. (2009). Training simulated patients: Evaluation of a training approach using self-assessment and peer/tutor feedback to improve performance. *BMC Medical Education*, 9(37), 1–6. <https://doi.org/10.1186/1472-6920-9-37>
- Pheister, M., Stagno, S., Cotes, R., Prabhakar, D., Mahr, F., Crowell, A., & Schwartz, A. (2016). Simulated Patients and Scenarios to Assess and Teach Psychiatry Residents. *Academic Psychiatry*, 114–117. <https://doi.org/10.1007/s40596-015-0477-8>
- Piot, M.-A., Dechartres, A., Attoe, C., Jollant, F., Lemogne, C., Layat Burn, C., Rethans, J.-J., Michelet, D., Cross, S., Billon, G., Guerrier, G., Tesniere, A., & Falissard,

- B. (2020). Simulation in psychiatry for medical doctors: A systematic review and meta-analysis. *Medical Education*, 54(8), 696–708. <https://doi.org/10.1111/medu.14166>
- Rakovshik, S. G., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. *Clinical Psychology Review*, 30(5), 496–516. <https://doi.org/10.1016/j.cpr.2010.03.004>
- Richards, D. (2011). Prevalence and clinical course of depression: A review. *Clinical Psychology Review*, 31(7), 1117–1125. <https://doi.org/10.1016/j.cpr.2011.07.004>
- Ritchie, H., & Roser, M. (2018). *Mental Health* [Published online at OurWorldInData.org]. <https://ourworldindata.org/mental-health>
- Smith, C. M., Gephardt, E. G., & Nestel, D. (2015). Applying the theory of Stanislavski to simulation: Stepping into role. *Clinical Simulation in Nursing*, 11(8), 361–367. <https://doi.org/10.1016/j.ecns.2015.04.001>
- Smith, C. M., Edlington, T. L., Lawton, R., & Nestel, D. (2015). The dramatic arts and simulated patient methodology. In D. Nestel and M. Bearman (Eds.), *Simulated Patient Methodology: Theory, Evidence and Practice* (pp. 39–45). John Wiley & Sons.
- Strauß, B., & Brähler, E. (2009). Forschungsgutachten zur Ausbildung von Psychologischen PsychotherapeutInnen und Kinder- und JugendlichenpsychotherapeutInnen. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie*, 59(5), 175–176. <https://doi.org/10.1055/s-0028-1090269>
- Thistlethwaite, J. E., & Ridgway, G. D. (2015). The content and process of simulated patient-based learning activities. In D. Nestel and M. Bearman (Eds.), *Simulated Patient Methodology: Theory, Evidence and Practice* (pp. 16–22). John Wiley & Sons.
- Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education*, 44(1), 85–93. <https://doi.org/10.1111/j.1365-2923.2009.03498.x>
- Wallace, J., Rao, R., & Haslam, R. (2002). Simulated patients and objective structured clinical examinations: Review of their use in medical education. *Advances in Psychiatric Treatment*, 8(5), 342–348. <https://doi.org/10.1192/apt.8.5.342>

- Weaver, M., & Erby, L. (2012). Standardized patients: A promising tool for health education and health promotion. *Health Promotion Practice, 13*(2), 169–174. <https://doi.org/10.1177/1524839911432006>
- Weck, F. (2013). *Psychotherapeutische Kompetenzen: Theorien, Erfassung, Förderung [Psychotherapy competencies: Theory, assessment, and training]*. Springer.
- Weck, F. (2017). Wie lässt sich die psychotherapeutische Ausbildung optimieren? *Zeitschrift Für Klinische Psychologie Und Psychotherapie, 46*, 71–72. <https://doi.org/10.1026/1616-3443/a000419>
- Wind, L. A., Van Dalen, J., Muijtjens, A. M. M., & Rethans, J.-J. (2004). Assessing simulated patients in an educational setting: The MaSP (Maastricht Assessment of Simulated Patients). *Medical Education, 38*(1), 39–44. <https://doi.org/10.1111/j.1365-2923.2004.01686.x>
- Wünderlich, M. S., Nissen, C., Feige, B., Philipsen, A. S., & Voderholzer, U. (2012). Portrayal of psychiatric disorders: Are simulated patients authentic? *Academic Psychiatry, 36*(6), 501–502. <https://doi.org/10.1176/appi.ap.11090163>
- Wünderlich, M. S., Peters, J., Philipsen, A., Kopasz, M., Berger, M., & Voderholzer, U. (2008). Einsatz von Simulationspatienten in den Lehrfächern Psychiatrie und Psychotherapie. *Der Nervenarzt, 79*(11), 1273–1282. <https://doi.org/10.1007/s00115-008-2493-3>
- Zendejas, B., Brydges, R., Wang, A. T., & Cook, D. A. (2013). Patient outcomes in simulation-based medical education: A systematic review. *Journal of General Internal Medicine, 28*(8), 1078–1089. <https://doi.org/10.1007/s11606-012-2264-5>

## 7. Appendix

### 7.1 Paper I

**Can simulated patient encounters appear authentic?  
Development and pilot results of a rating instrument based on the  
portrayal of depressive patients**

Destina Sevde Ay-Bryson<sup>a</sup>, Florian Weck and Franziska Kühne

Clinical Psychology and Psychotherapy, University of Potsdam, Germany

<sup>a</sup>Correspondence concerning this article should be addressed to Destina Ay-Bryson, Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany, Email address: destina.ay@uni-potsdam.de

#### **Copyright statement**

This is a raw version of the manuscript that was accepted for publication in “Training and Education in Professional Psychology”:

Ay-Bryson, D. S., Weck, F., Kühne, F. (2020). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000349>

**Acknowledgements**

We wish to express our gratitude to Tatjana Paunov for testing and providing feedback on the initial version of APD, special thanks to Samuel Oliver Ay-Bryson for being part of the back-translation of the APD and NESP, to our interns Lena Walter and Elisa Kozlowski for their support during the process of preparing the study, and to Dr. Brian Bloch for his comprehensive proofreading. Finally, we thoroughly appreciate the comments provided throughout the review process.

**Disclosure**

On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Funding**

The Hans Böckler Foundation supported this research with a full PhD scholarship to Destina S. Ay-Bryson.

### Abstract

**Objective.** For training purposes, simulated patients (SPs), i.e., healthy people portraying a disorder, are disseminating more into clinical psychology and psychotherapy. In the current study we developed an observer-based rating instrument for the evaluation of SP authenticity, namely it not being possible to distinguish them from real patients, so as to foster their use in evidence-based training. **Methods.** We applied a multi-step inductive approach to develop the *Authenticity of Patient Demonstrations* (APD) scale.  $N = 97$  independent psychotherapy trainees, 77.32% female, mean age of 31.49 ( $SD = 5.17$ ) years, evaluated the authenticity of two independent SPs, each of whom portrayed a depressive patient. **Results.** The APD demonstrated good internal consistency (Cronbach's  $\alpha = .83$ ) and a strong correlation ( $r = .82$ ) with an established tool for assessing SP performance in medical contexts. The APD scale distinguished significantly between an authentic and unauthentic SP ( $d = 2.35$ ). **Conclusions.** Preliminary evidence for the psychometric properties of the APD indicates that the APD could be a viable tool for recruiting, training, and evaluating the authenticity of SPs. Strengths, limitations and future directions are also discussed in detail.

**Keywords:** authenticity; evidence-based training; simulated patients; role-play; mental disorders

### Significance statement

This study demonstrates that simulated patients (SPs) can authentically portray a depressive case. The results provide preliminary evidence of psychometrically sound properties of the rating scale that contributes to distinguishing between authentic and unauthentic SPs, and may thus foster SPs' dissemination into evidence-based training.

### Introduction

While cognitive behavioral therapy (CBT) has repeatedly been demonstrated to be an effective treatment for a wide range of mental health problems (Butler, Chapman, Forman, & Beck, 2006; Lambert, 2013), the systematic evaluation of psychotherapy training is still in its infancy and is currently not considered to be evidence-based (Campbell, Norcross, Vasquez, & Kaslow, 2013; Fairburn & Cooper, 2011; Rakovshik & McManus, 2010). One way to train and evaluate psychotherapists in training is the use of role-play interactions with simulated patients or standardized patients (SPs; e.g. Miller, 1990; Muse & McManus, 2013). Simulated patients are healthy laypersons who (repeatedly) simulate a bodily disease or mental health issue (Barrows, 1993). SPs may provide learners with the opportunity to learn and practice in a controlled environment that learners may experience with increasing confidence (Pheister et al., 2016). They may also provide vital feedback to the learner. Due to their many benefits, such as availability, repeatability and flexibility, the use of SPs is increasing in popularity. At the same time, however, this method is associated with many resources, such as high costs and, importantly, issues concerning the convincing portrayal of symptoms (Barrows, 1993; Kühne, Ay, Otterbeck, & Weck, 2018). Consequently, to demonstrate the training effects and external validity of this method, the *authenticity* of SPs becomes crucial. As Wind et al. (2004) put it: “For teaching purposes the authenticity of role play is more important than consistency [...]” (p. 39).

By authenticity we mean the “impossibility of distinguishing SPs from (real) patients” (Wünderlich, Nissen, Feige, Philipsen, & Voderholzer, 2012; p. 501). Hence, symptoms portrayed by SPs should be indistinguishable from those portrayed by a real patient. One pilot-study (Wünderlich et al., 2008) investigated whether SPs can portray mental health problems realistically. According to the authors, study participants were surprised by the

SPs' authenticity, although no statistical analysis was reported. The authenticity of SPs is regarded as particularly relevant, because it strongly regulates and impacts on learning experience (Wünderich et al., 2012). In a study conducted by Krahn et al. (2001), learners experienced the SPs as less authentic than real patients, which led to difficulties among the learners in terms of feeling empathetic, a therapeutic skill considered highly important. Hence, following the results reported by Krahn et al. (2001), it is crucial to provide authentic SPs, so as to enable learners to acquire highly important therapeutic skills.

Many studies address the authenticity of SPs (Kühne et al., 2018), for instance, that it may be challenging for SPs to portray symptoms in a realistic and authentic manner. However, little to no empirical data is available to draw firm conclusions as to *how* authenticity may be positively influenced, or even trained. Further, most publications are conducted in medical or nursing contexts, as SPs have been regarded with skepticism in psychology-related subjects (Hodges, Hollenberg, McNaughton, Hanson, & Regehr, 2014). However, this is currently changing. Importantly, a pending reform of the psychotherapy law in Germany (Deutscher Bundestag, 2019) prescribes university examinations of psychotherapy students using SPs, similar to the Objective Structure Clinical Examination in medical education. Hence, with respect to an increasing use of SPs in psychotherapy training, it is necessary to provide insights into how authenticity can be improved. However, for this to be investigated systematically, a reliable and valid instrument is needed to measure authenticity.

To the best of our knowledge, no instrument so far has been designed to assess the authenticity of SPs, especially for psychotherapeutic contexts, which we intended to target. The newly developed instrument was intended to be an easy-to-use rating tool that can be used regardless of the (mental) disorder to be demonstrated.



Hence, the first aim of the present study was to pilot-validate the observer-based rating-instrument, the *Authenticity of Patient Demonstrations* (APD) scale, based on the portrayal of a depressive patient. Regarding the reliability of the APD scale, we expected the instrument to demonstrate internal consistency and medium to strong item-total correlations. With regard to the validity of the APD scale, we hypothesized that the APD would distinguish between an authentic and an unauthentic SP. Further, regarding the convergent validity, we expected strong correlations between the APD scale and depressive symptoms, as our SPs were to portray a depressive patient. Consequently, we expected small correlations with anxiety symptoms with respect to discriminant validity. Moreover, we aimed to explore the factor structure of the APD.

Our second aim was to explore effects of the current raters. Rating methods are widespread (Wirtz, 2017), but, few authors report whether observational effects occur. Since the raters of the current study judged two SP videos successively, we explored whether there was an order effect between the two conditions of our study design, namely whether raters judged the SP as even more authentic when they first saw an unauthentic SP, in comparison to when they saw the authentic SP first, and vice versa.

## **Methods**

### **Development of the *Authenticity of Patient Demonstrations* (APD)**

To assess the authenticity of SPs, we developed the observer-based rating scale *Authenticity of Patient Demonstrations* (APD; original version in German: *Authentizität von Patientendarstellungen*), for which we conducted a multi-step development and evaluation. Our first aim was an inductive search for the most appropriate items for assessing the authenticity of patient portrayals. An inductive approach is mainly applied when there is no one main theory underlying the item construction (Burisch, 1984). Although pre-existing instruments, such as the Nijmegen Evaluation of the Simulated

Patient (NESP; Bouter, van Weel-Baumgarten, & Bolhuis, 2013) and the Maastricht Assessment of Simulated Patients (MaSP; Wind et al., 2004), for assessing SP performances were helpful as a first source of information, they did not fully conform to our objectives; neither measure is specifically tailored for psychotherapeutic contexts. Moreover, both assess SPs in the medical context without blinding the rater as to whether the person shown is an SP or a real patient. Therefore, we employed several sources of information in order to develop our items inductively. Constantly considering the NESP and MaSP, the authors DAB and FK gathered information based on an SP-manual (Peters & Thrien, 2018), and noted potentially relevant items based on ten pre-existing videos of role-play interactions between psychology students, as well as based on two randomly selected videos of a psychotherapy-teaching DVD (Brakemeier & Jacobi, 2017). Second, based on these notes and information, we formulated 15 initial items, whereby we incorporated two negatively worded items. These items were then tested by DAB, FK and a student researcher (B.Sc. psych.) who checked whether the items were complete and whether they needed to be reworded for better comprehensibility. Because we intended to develop an instrument that could be used regardless of patient diagnosis and that does not disclose the SP status per se, we used the term *person* instead of SP in the items. After adjustments were made, one expert in psychotherapy, FW, used the items on two further videos from the training DVD (Brakemeier & Jacobi, 2017) and provided feedback. This resulted in further adjustments, namely eliminating and rewording items. The items utilize a 4-point Likert scale: 0 = “strongly disagree”, 1 = “disagree”, 2 = “agree”, 3 = “strongly agree”.

### ***Content validity and applicability***

Third, we conducted an online expert survey with psychologists and associates of theater arts to examine the content validity (i.e. relevance and comprehensibility) of the

APD, which at that point contained 11 items. In order to give experts the opportunity to express uncertainty and thus to provide feedback for items that ought to be improved, we used a 5-point Likert scale, i.e. 1 = “not very relevant/comprehensible” to 5 = “very relevant/comprehensible”. Additionally, free comment fields were provided for each of the items of the APD, and at the end, we asked the experts to note whether any aspect of patient portrayal authenticity had been omitted. We conducted the survey in UP Survey, which is a service of the University of Potsdam. Finally, the APD was revised according to the results of the online survey; see Table 1. The final and currently evaluated version of the APD with ten items is presented in Table 2.

### ***Translation into English***

While the APD was pilot-validated within a German-speaking sample, we translated the measure into English to encourage further validation studies in English-speaking samples. To translate the items, we used a method similar to the dual panel method (Hagell, Hedin, Meads, Nyberg, & McKenna, 2010). An English native speaker and translator translated the original version of the APD into the English language. One co-author (FK) independently checked overlap between the two versions and highlighted discrepancies. This was followed by a panel including DAB, FW and FK (developers of the APD), and two other psychotherapy researchers who are fluent in English. Subsequently, DAB made final adjustments to the translated items, which were then finally approved by the group. The translated items are presented in Table 2.

### **Video material**

In order to investigate the reliability, dimensionality as well as the validity of the APD scale, we produced three video tapes, all featuring a three-minute therapy segment (behavioral activation) with three different SPs. The SPs were psychology master students, one of whom has had substantially more experience in portraying a depressive

patient due to her being a staff member of our Department, while one of the other two was an intern and the other a staff member at the outpatient clinic. All SPs received the same instruction, which translates as follows: “Please carefully read the following role-script and try to empathize with the role. Consider what you, if you were this depressive patient, would say during the patient-therapist interaction. Afterwards you will be asked to simulate the depressive patient. You have fifteen minutes to prepare yourself”. These portrayals were then presented to two experts (licensed psychotherapists) who ranked the SPs from *most authentic* to *least authentic*; the rankings coincided 100%. Consequently, we selected the video with the highest ranking as the “authentic case” and the video with the lowest ranking as the “unauthentic case” for our study. The SPs portrayed a depressive patient by reporting on their normal daily activities in response to the therapist’s request: “Please tell me what a regular day looks like for you.”. Specifically, the two SPs’ authenticity differed with regard to their way of speaking (i.e., loud vs. quiet, fast vs. slowly), posture (i.e., open vs. closed) and disorder-specific content (i.e., reporting on low vs. high levels of physical activity during their daily lives). Each participant watched both videos, whereby the order of the videos alternated (Figure 1).

### **Online study and sample**

To pilot-validate the APD, we used a within-subjects design. Participants were informed about the content of the study and then alternately allocated to one of two conditions, see Figure 1. They were further informed that they would be shown two videos for rating. However, they were not informed about the order and the different conditions of the videos. While in condition A, the authentic case was presented first (t1), followed by the unauthentic case (t2), in condition B the order was reversed. The study was conducted online in UP Survey. Participants were recruited via several email lists and newsletters and did not receive any financial reimbursement. Each participant

provided informed consent prior to participation. Our university's ethics commission approved the study (no. 1/2019, University of Potsdam).

Psychotherapists in training or licensed psychotherapists were eligible to participate. A total of 101 participants (from now on referred to as "raters") participated in the study. However, four raters had to be excluded for technical issues (e.g. video did not play;  $n = 2$ ), one outlier and because one rater did not answer the sociodemographic questions, such as age and gender. Thus, a total sample of  $N = 97$  (condition A = 48, condition B = 49) could be included in the analyses (for the sociodemographic data, see Table 3). Of the psychotherapists in training, fourteen (16.67%) were in their first year of psychotherapy training, 19 (22.62) were in year two, 24 (28.57%) in year three, 17 (20.24%) in year four, 7 (8.3%) in year five, and one (1.19%) each in year six, seven and eight.

## Measures

### *Nijmegen Evaluation of the Simulated Patient (NESP)*

To evaluate the convergent validity of the APD, we translated the Nijmegen Evaluation of the Simulated Patient (NESP; Bouter et al., 2013) into German. The NESP is a standardized, three-factorial instrument that is used in various medical contexts to assess SP performances. The three components of the NESP include a) role-play, b) process of feedback, and c) application of feedback rules. In the current study, we used the 9-item subscale "role play" (Cronbach's  $\alpha = .86$ ). This subscale assesses an SP's acting ability (example item: "The SP knew his or her role well.") and their ability to adjust to the student's level during role-play (example item: "The SP adjusted the role naturally to the level of the student."). A dichotomous yes/no answer-format was applied. During the translation process we followed the steps suggested by Gudmundsson (2009), applying the back-translation method. In short, DAB translated the original version of the NESP into German, which was then back-translated into English by an English native

speaker and translator. Overlap and discrepancies between the back-translated version and the original were discussed between DAB and FK. Finally, the German version was compared with the original one including a third researcher (FW). The German NESP can be obtained from DAB upon request.

### ***Brief Symptom Inventory (BSI)***

We applied the 6-item subscales on depression and anxiety of the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983; German version: Franke, 2000) to assess the convergent and discriminant validity of the APD, respectively. The BSI is a self-report questionnaire designed to measure various levels of psychopathology. A 5-point scale ranging from 0 = “not at all” to 4 = “extremely” was used. We used an item version reworded for observers. Previous studies revealed the BSI to have good reliability and validity (Boulet & Boss, 1991; Broday & Mason, 1991; Franke, 2000). In the current study Cronbach’s  $\alpha$  was .79 for the depression subscale, and  $\alpha = .81$  for the anxiety subscale.

### **Additional items**

We additionally implemented one qualitative and three quantitative items, for which we referred to Wünderlich et al. (2012). The qualitative item assessed the assumed diagnosis: “*In your opinion, which diagnosis is most likely?*”

The three quantitative items were rated on an 11-point scale ranging from 0% to 100%: “How likely is it that the person in the video is a real patient?” (item *real*), “How likely is it that the person in the video is depressive?” (item *depressive*), and “How likely is it that the person in the video is anxious?” (item *anxious*).

### **Statistical analyses**

#### ***Reliability***

Internal consistency was calculated with Cronbach's  $\alpha$ , whereby a value of .70 to .80 is regarded as "satisfactory". In addition, item-total correlations were calculated, whereby correlations from .40 to .70 are regarded as "good" (Kevala & Moosbrugger, 2012).

### ***Validity and dimensionality***

In order to check the APD's content validity, we computed means, standard deviations and frequencies of expert comments regarding comprehensibility and relevance. To check the validity of the APD we first computed a paired  $t$ -test to analyze mean differences between authentic and unauthentic APD mean scores. Pearson's correlation coefficients were calculated to demonstrate convergent and discriminant validity. Moreover, we checked the qualitative item (*assumed diagnosis*) as an indicator of validity. With regard to the dimensionality, APD mean scores were subjected to exploratory factor analyses (EFA) applying the principal axes factor analysis for oblique rotation ( $\delta = 0$ , oblimin). In preparation for the EFA, we investigated the skewness and kurtosis of the variables. To test the suitability of the data for the EFA, we computed the Kaiser-Meyer-Olkin measure. We referred to the Kaiser-1 heuristic (factors corresponding to eigenvalues  $> 1$ ) as well as the parallel analysis (comparison of randomly produced eigenvalues; O'Connor, 2000) in order to determine the number of factors. We used the R package *paran* (Dinno & Dinno, 2018) to compute the parallel analyses.

### ***Order effect of presentation***

We computed  $t$ -tests to determine whether there was a significant difference between having observed the authentic or the unauthentic SP first. All statistical analyses were performed with R version 3.4.2 (2017); the alpha level was set to  $\alpha = .05$ .

## **Results**

### **Content validity and applicability**

Ten experts participated in the survey; three were licensed psychotherapists, one was a psychotherapist in training, two were researchers in psychology, and four were associates of theater arts. The majority (80%) of participants was female, and the mean age was 34.56 years ( $SD = 7.06$ ). The average work experience in their corresponding field was 7.93 years ( $SD = 5.10$ ). Results for the content validity are presented in Table 1. Two items were perceived as less comprehensible and relevant compared to the other items, which is why we eliminated both of them. In the open comment fields, the experts indicated improvements for the German phrasing and wording.

### **Descriptive results**

Table 2 shows item means, standard deviations and factor loadings for the authentic and unauthentic SP as well as over both SPs. On average, the raters *disagreed to agreed* that the unauthentic SP portrayed the patient authentically ( $M = 1.43$ ,  $SD = .52$ ). By contrast, the raters *agreed to strongly agreed* that the authentic SP did indeed portray the patient authentically ( $M = 2.50$ ,  $SD = .39$ ). The difference in APD mean scores between the authentic and unauthentic SP was significant ( $t(96) = 16.70$ ,  $p < .001$ ; Cohen's  $d = 2.35$ ).

### **Reliability**

Regarding the reliability of the APD scale, we found Cronbach's  $\alpha$  to be .83 (authentic SP: .85; unauthentic SP: .84), indicating good internal consistency. Most item-total correlations ranged from  $r = .40$  to  $.78$  ( $r_{\text{authentic}} = .41$  to  $.81$ ;  $r_{\text{unauthentic}} = .44$  to  $.80$ ), with the exception of item 6 ( $r = .27$ ,  $r_{\text{unauthentic}} = .18$ ) and item 7 ( $r_{\text{authentic}} = .30$ ).

### **Dimensionality**

With respect to the EFA, all variables were acceptable regarding skewness and kurtosis. The Kaiser-Meyer-Olkin measure ( $KMO = .86$ ) indicated the suitability of the data for the EFA. According to the Kaiser-1 heuristic, one factor with an eigenvalue of



3.86 emerged (authentic case: 4.13; unauthentic case: 4.01). Further, results of the parallel analysis (see Supplementary Material 1) indicated a one-factorial structure explaining 38% of the variance (authentic: 41%; unauthentic: 40%). All factor loadings are presented in Table 2.

### **Convergent validity**

Regarding convergent validity, we revealed a strong positive correlation between the APD and NESP means ( $r = .82, p < .001$ ), and a small positive, but non-significant, correlation between the APD and the BSI depression subscale  $r = .16, p = .11$  ( $r_{\text{authentic}} = .19, p = .06; r_{\text{unauthentic}} = .19, p = .06$ ). A medium to strong positive correlation between the APD and the item *real*  $r = .43, p < .001$  ( $r_{\text{authentic}} = .46, p < .001; r_{\text{unauthentic}} = .55, p < .001$ ), as well as the item *depressive*  $r = .50, p < .001$  ( $r_{\text{authentic}} = .45, p < .001; r_{\text{unauthentic}} = .55, p < .001$ ) was shown. In this context, the raters indicated in the qualitative item that a depressive disorder is most likely for the authentic SP (100% consensus).

### **Discriminant validity**

With respect to discriminant validity we found no correlation between the APD and the BSI anxiety subscale  $r = .004, p = .97$  ( $r_{\text{authentic}} = -.07, p = .49; r_{\text{unauthentic}} = .05, p = .63$ ). Furthermore, we found no correlation between the APD and the item *anxious* over both samples ( $r = .07, p = .52$ ), but a small negative and a small positive correlation for the authentic and unauthentic SP, respectively ( $r_{\text{authentic}} = -.22, p < .05; r_{\text{unauthentic}} = .25, p < .05$ ). Finally, regarding the qualitative item for the unauthentic SP, there was less agreement: 20.62% found another disorder (e.g., narcissistic personality disorder or adjustment disorder) or no disorder at all more likely.

### **Order effect of presentation**

The raters judged the authentic SP as even more authentic ( $M = 2.65, SD = .34$ ) when they first saw the unauthentic SP (condition B) than when the authentic SP was seen first

(condition A;  $M = 2.35$ ,  $SD = .38$ ;  $t(93.28) = -4.19$ ,  $p < .001$ ; Cohen's  $d = -.85$ ). Similarly, the unauthentic SP was rated even less authentic when raters were first presented the authentic SP (condition A;  $M = 1.31$ ,  $SD = .55$ ) compared to when the unauthentic SP was seen first (condition B;  $M = 1.54$ ,  $SD = .46$ ;  $t(91.60) = -2.28$ ,  $p < .05$ ; Cohen's  $d = -.47$ ). These results indicate an order effect (see Figure 2).

### Discussion

Our objective was to develop and pilot-validate a tool for independent raters that can easily be used to assess the authenticity of SPs within clinical psychology and psychotherapeutic contexts. We have reported the multi-step development and evaluation of our observer-based rating scale and provide a translation into English. Moreover, we conducted a detailed survey with respect to the APD's content validity. Altogether, the current study provides preliminary evidence of psychometrically sound properties of the APD, based on the portrayal of a depressive patient. As hypothesized, the scale demonstrated good internal consistency, good item-total correlations, and a one-factor solution of the APD was suggested. Further, as expected, we have shown that the APD scale distinguished well between an authentic and unauthentic SP. Finally, we found indices for convergent and discriminant validity of the APD scale.

Overall, the results are comparable for the authentic and unauthentic SP, which leads us to conclude that the APD may be used for authentic as well as unauthentic portrayals. As expected, the APD demonstrated good internal consistency, which is comparable with other measures for assessing SP performances from medical education contexts (NESP:  $\alpha = .92$ ; MaSP:  $\alpha = .73$ ).

Moreover, the EFA suggested a one-factor solution. Although the factor analysis was based on a small sample size, we consider our results as interpretable, albeit with caution (MacCallum, Widaman, Zhang, & Hong, 1999). Item 7 demonstrated a lower factor

loading than the other items for the authentic SP. Similarly, item 6 demonstrated a lower factor loading for the unauthentic case. However, we refrained from extracting an additional factor, as eigenvalues of the EFA did not suggest a need for this.

Moreover, the strong positive correlation between the APD and the NESP (Bouter et al., 2013) is plausible, as the NESP was one main source of information for the item development of the APD. Contrary to our expectation, we did not find a strong correlation between the APD and BSI depression subscale (Franke, 2000). This may be due to the fact that the items of the BSI focus directly on explicit symptoms of depression, e.g., suicidal thoughts. While our SPs were instructed to demonstrate a depressive patient, not all symptoms were explicitly incorporated into the 3-minute portrayals, which represents a limitation of the current study. Henceforth, script planning for SP interactions is crucial. On the other hand, we found strong correlations between the APD and the items that referred to a general clinical impression of the SP as a *real* and *depressive* patient, which may be considered indicative of convergent validity. As expected, we found no correlation between the APD and the BSI anxiety subscale. Surprisingly, we found a small negative correlation between the APD and the additional item *anxious* for the authentic SP. This suggests that raters did not perceive the authentic SP as anxious, and that they rated the SP with more confidence accordingly. By contrast, we revealed a small positive correlation between the APD and the item *anxious* for the unauthentic SP. Arguably, the raters perceived the unauthentic SP as generally more anxious, which may in turn be associated with inauthenticity. However, this should be interpreted with caution, as further investigation on the interrater reliability is required.

On a more explorative note, we investigated whether there was a specific rater effect, namely an order effect. As the results have shown, the authentic SP was perceived even more authentic when raters first saw an unauthentic SP and vice versa. Although there is

only sparse data so far on order effects in rating studies, this result is highly plausible. Based on our data, we would recommend studies to use rater data to examine and report such and similar effects, such as halo, leniency or strictness effects (Wirtz, 2017).

Despite promising preliminary results of the current study, some limitations need to be discussed. First, in the present study, only SPs could be considered. While ethical reasons hindered a comparison between SPs and real patients, as the current study was conducted online, which also made online distribution of the video materials inevitable, it is crucial for future research to provide such a comparison, as the definition of *authenticity* relies on this very comparison. Second, only the portrayal of a depressive patient was considered in this study. While we demonstrated that it is indeed possible to portray a depressive patient authentically, the current results need replication with SPs portraying other mental health problems. Further, although the current study design has proven to be a feasible and cost-effective design for pilot-validation, we included a large number of raters who evaluated two SPs, as opposed to commonly used empirical study designs. Consequently, we could not provide an analysis of interrater reliabilities. Hence, follow-up studies should consider designs to investigate the interrater reliabilities of the APD. Finally, in the current study, participants were blind to the status of the SPs regarding their being SPs, as well as the diagnosis they were instructed to portray. While this is a strength of the present study, we cannot conclude that the low authenticity ratings of the unauthentic SP are due to the raters' assumption that the SP did a poor job in depicting a *depressive* patient. Possibly, the unauthentic SP was believed to be portraying something else, which is why the SP was given a low rating. This could be further investigated by instructing raters accordingly.

**Implications**

In sum, we provided a scale development of a tool with the potential to improve the authenticity of SPs, which in turn may improve evidence-based training of psychotherapists. As the present study delivers preliminary results on this tool, future studies are encouraged to investigate (a) a comparison between SPs and real patients, (b) the authenticity of portrayal of mental health problems other than depression, and (c) interrater reliabilities of the APD as discussed above. As for the next step, it would be desirable to explore which factors, such as personality traits or prior experience with mental health problems, influence SP authenticity (Wünderich et al., 2012).

### References

- Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Academic Medicine*, 68(6), 443–51.
- Boulet, J., & Boss, M. W. (1991). Reliability and validity of the Brief Symptom Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(3), 433.
- Bouter, S., van Weel-Baumgarten, E., & Bolhuis, S. (2013). Construction and validation of the nijmegen evaluation of the simulated patient (nesp): assessing simulated patients' ability to role-play and provide feedback to students. *Academic Medicine*, 88(2), 253–259.
- Brakemeier, E.-L., & Jacobi, F. (2017). *Verhaltenstherapie in der Praxis [Film]*. Weinheim: Beltz.
- Broday, S. F., & Mason, J. L. (1991). Internal consistency of the Brief Symptom Inventory for counseling-center clients. *Psychological Reports*, 68(1), 94–94.
- Burisch, M. (1984). Approaches to personality inventory construction: a comparison of merits. *American Psychologist*, 39(3), 214.
- Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical Psychology Review*, 26(1), 17–31.
- Campbell, L. F., Norcross, J. C., Vasquez, M. J., & Kaslow, N. J. (2013). Recognition of psychotherapy effectiveness: The APA resolution. *Psychotherapy*, 50(1), 98.
- Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: an introductory report. *Psychological Medicine*, 13(3), 595–605.
- Derogatis, L. R., & Spencer, P. M. (1993). *Brief symptom inventory: BSI* (Vol. 18). Pearson Upper Saddle River, NJ.

- Deutscher Bundestag. *Entwurf eines Gesetzes zur Reform der Psychotherapeutenausbildung*. , Pub. L. No. 19/9770 (2019).  
[https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/Gesetze\\_und\\_Verordnungen/GuV/P/Psychotherapeutenausbildung\\_Reform\\_Bundestag.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/P/Psychotherapeutenausbildung_Reform_Bundestag.pdf)
- Dinno, A., & Dinno, M. A. (2018). *Package 'paran.'*
- Fairburn, C. G., & Cooper, Z. (2011). Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy*, 49(6–7), 373–378.
- Franke, G. H. (2000). BSI, brief symptom inventory, von LR Derogatis-Deutsches manual. *Göttingen: Beltz Test Gesellschaft*.
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, 61(2), 29–45.
- Hagell, P., Hedin, P.-J., Meads, D. M., Nyberg, L., & McKenna, S. P. (2010). Effects of method of translation of patient-reported health outcome questionnaires: a randomized study of the translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden. *Value in Health*, 13(4), 424–430.
- Hodges, B. D., Hollenberg, E., McNaughton, N., Hanson, M. D., & Regehr, G. (2014). The psychiatry OSCE: a 20-year retrospective. *Academic Psychiatry*, 38(1), 26–34.
- Kevala, A., & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilung. *Testtheorie Und Fragebogenkonstruktion*, 2, 75–102.
- Krahn, L. E., Sutor, B., & Bostwick, J. M. (2001). Conveying emotional realism: a challenge to using standardized patients. *Academic Medicine*, 76(3), 216–217.

- Kühne, F., Ay, D. S., Otterbeck, M., & Weck, F. (2018). Standardized patients in clinical psychology and psychotherapy: A scoping review on barriers and facilitators for implementation. *Academic Psychiatry, 6*, 773–781.
- Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, 6*, 169–218.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), S63–7.
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review, 33*(3), 484–499.
- Peters, T., & Thrien, C. (2018). *Simulationspatienten: Handbuch für die Aus- und Weiterbildung in medizinischen- und Gesundheitsberufen*. Hogrefe.
- Pheister, M., Stagno, S., Cotes, R., Prabhakar, D., Mahr, F., Crowell, A., & Schwartz, A. (2016). Simulated Patients and Scenarios to Assess and Teach Psychiatry Residents. *Academic Psychiatry, 1*–4.
- Rakovshik, S. G., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. *Clinical Psychology Review, 30*(5), 496–516.
- Wind, L. A., Van Dalen, J., Muijtjens, A. M., & Rethans, J. J. (2004). Assessing simulated patients in an educational setting: the MaSP (Maastricht Assessment of Simulated Patients). *Medical Education, 38*(1), 39–44.



- Wünderich, M., Peters, J., Philipsen, A., Kopasz, M., Berger, M., & Voderholzer, U. (2008). Einsatz von Simulationspatienten in den Lehrfächern Psychiatrie und Psychotherapie. *Der Nervenarzt*, *79*(11), 1273–1282.
- Wünderich, M. S., Nissen, C., Feige, B., Philipsen, A. S., & Voderholzer, U. (2012). Portrayal of psychiatric disorders: Are simulated patients authentic? *Academic Psychiatry*, *36*, 501–502.

**Table 1***Content validity results for preliminary items of the APD (N = 10)*

<b>Item number</b>	<b>Comprehensibility</b> <i>M (SD)</i>	<b>Relevance</b> <i>M (SD)</i>	<b>Free comment</b> Frequency <sup>a</sup>	<b>Action<sup>b</sup></b>
<b>1.</b>	4.2 (1.08)	4.8 (0.4)	2	reworded
<b>2.</b>	2.6 (1.28)	3.0 (1.4)	8	eliminated
<b>3.</b>	4.4 (0.92)	4.0 (0.89)	2	reworded
<b>4.</b>	4.1 (0.94)	4.6 (0.66)	4	split into two
<b>5.</b>	4.6 (0.66)	4.4 (0.92)	1	unchanged
<b>6.</b>	3.3 (1.1)	4 (1)	4	reworded
<b>7.</b>	4.2 (0.98)	4.2 (0.87)	1	unchanged
<b>8.</b>	4.5 (0.67)	4.3 (0.9)	3	reworded
<b>9.</b>	3.5 (1.03)	3.2 (1.08)	7	eliminated
<b>10.</b>	3.7 (1.35)	3.6 (1.2)	5	reworded
<b>11.</b>	4 (1)	3.4 (1.2)	5	reworded

*Note.* A 5-point Likert scale was used: 1 = not very relevant/comprehensible to 5 = very relevant/comprehensible); <sup>a</sup> frequency = number of comments made by experts; <sup>b</sup> action describes how we proceeded with the item.

**Table 2***Final items of the APD, descriptive results and factor loadings (N = 97)*

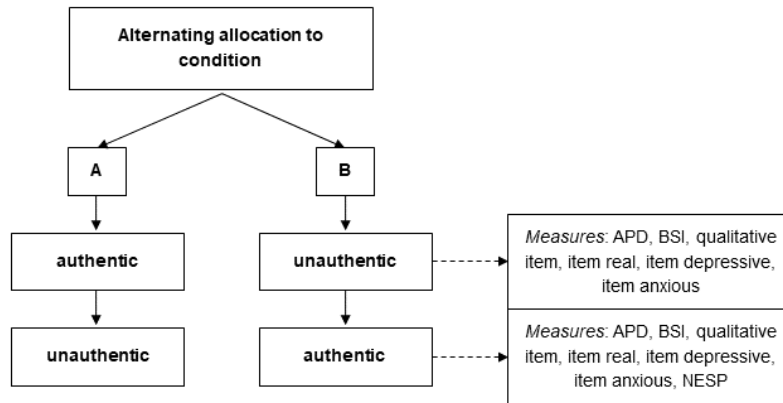
Item	Total		Authentic SP		Unauthentic SP	
	M (SD)	FL	M (SD)	FL	M (SD)	FL
1. Overall, the person describes his/her disorder convincingly. (Die Person beschreibt ihre Erkrankung insgesamt überzeugend.)	1.88 (.46)	.72	2.54 (.52)	.73	1.23 (.73)	.78
2. The posture is appropriate for the disorder. (Die Körperhaltung ist der Erkrankung angemessen.)	2.11 (.51)	.66	2.54 (.56)	.72	1.68 (.73)	.53
3. The person's facial expressions are appropriate for the disorder. (Die Mimik der Person ist dem Störungsbild angemessen.)	1.86 (.53)	.80	2.61 (.55)	.73	1.11 (.85)	.81
4. The person's gestures are appropriate for the disorder. (Die Gestik der Person ist dem Störungsbild angemessen.)	2.06 (.49)	.69	2.48 (.56)	.73	1.63 (.77)	.57
5. The way of speaking (pitch, tempo) is appropriate for the disorder. (Die Art und Weise des Sprechens (Tonlage, Geschwindigkeit) ist der Erkrankung angemessen.)	1.66 (.54)	.76	2.36 (.65)	.80	.96 (.83)	.77
6. The person maintains his/her perspective as a patient during the entire interaction. (Die Person bleibt die gesamte Interaktion über in ihrer Perspektive als Patient.)	2.21 (.53)	.24	2.51 (.60)	.55	1.92 (.83)	.17
7R. The person portrays the disorder unbelievably. (Die Person stellt das Störungsbild unglaubwürdig dar.)	2.10 (.59)	.50	2.59 (.75)	.29	1.57 (.85)	.68
8. The person describes disorder-specific symptoms appropriately. (Die Person beschreibt störungsspezifische Symptome angemessen.)	2.03 (.44)	.44	2.59 (.50)	.56	1.47 (.69)	.42
9. The person's behavior is appropriate for the disorder. (Das Verhalten der Person ist dem Erkrankungsbild angemessen.)	1.93 (.44)	.73	2.53 (.50)	.72	1.33 (.73)	.81
10R. The person responds as if he/she had learned the answers by heart beforehand. (Die Person antwortet, als hätte sie die Antworten vorab auswendig gelernt.)	1.82 (.66)	.39	2.27 (.78)	.39	1.37 (.95)	.47
<b>Total</b>	1.96 (.33)	-	2.50 (.39)	-	1.43 (.52)	-

*Note.* Brackets = German Version; R = Recode; SP = Simulated Patient; M = Mean; SD = Standard Deviation; FL = Factor Loadings; A 4-point Likert scale was used: 0 = “strongly disagree”, 1 = “disagree”, 2 = “agree”, 3 = “strongly agree”.

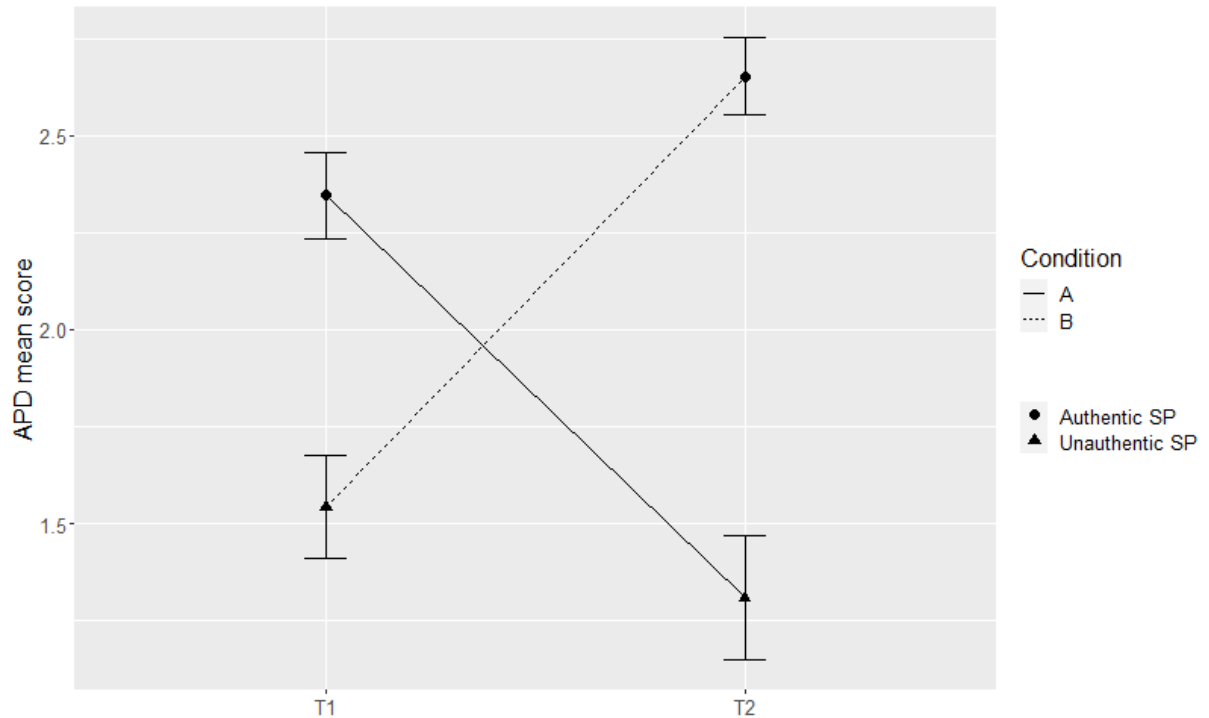
**Table 3***Sociodemographic data of the raters*

	<i>n</i>	<i>%</i>	<i>M</i> years ( <i>SD</i> )	<i>Range</i>
Age	97	n/a	31.49 (5.17)	24 – 48
Female	75	77.32	n/a	n/a
Studied psychology	80	82.47	n/a	n/a
Specialized in CBT	89	91.75	n/a	n/a
Clinical experience	n/a	n/a	2.64 (1.83)	0 – 8.5
<b>Current status</b>				
Licensed psychotherapist	12	12.37	n/a	n/a
Psychotherapist in training	85	87.63	n/a	n/a
<b>Training stage<sup>a</sup></b>				
		<b>Completed</b>	<b>Currently attending</b>	<b>Pending</b>
		<i>n</i> (%)		
Year 1 (“Psychiatry”)	84 <sup>b</sup>	58 (69.05)	17 (20.24)	9 (10.71)
Year 2 (“Psychosomatics”)	84	55 (65.48)	15 (17.86)	14 (16.67)
Year 3 (“Practical training”)	84	8 (9.52)	51 (60.71)	25 (29.76)
Theory	84	30 (35.71)	52 (61.91)	2 (2.38)

*Note.* <sup>a</sup> = Prototypically, psychotherapy training in Germany takes three to five years consisting of five training stages, whose order is flexible. *Year one* of training contains a 1200 hour-placement at a psychiatry, *year two* involves a 600 hour-placement at a psychosomatic facility, and in *year three* trainees start treating outpatients for at least 600 hours accompanied by regular supervision. Throughout the training, trainees receive 600 hours of *theory* courses; <sup>b</sup> = One psychotherapist in training did not answer questions to current trainings status; n/a = not applicable.

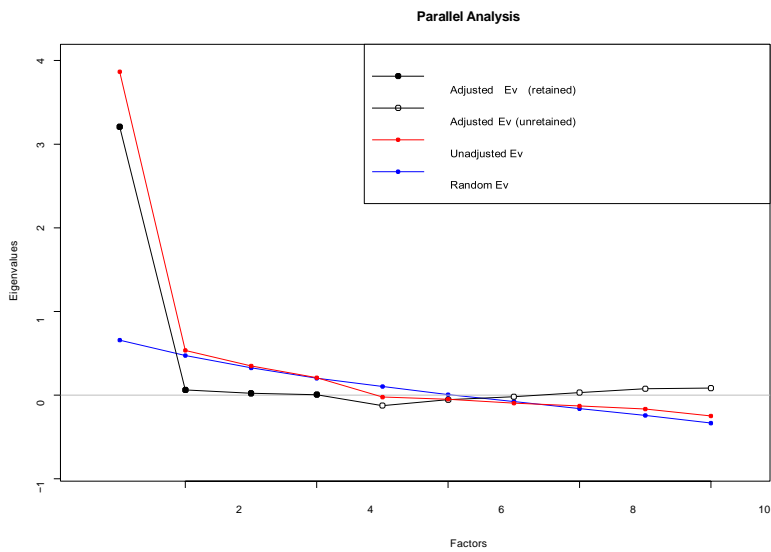
**Figure 1***Study design*

*Note.* APD = Authenticity of Patient Demonstrations; BSI = Brief Symptom Inventory; NESP = Nijmegen Evaluation of the Simulated Patient; the NESP was not administered at t1 to prevent bias at t2 since participants may have been triggered that the person is an SP.

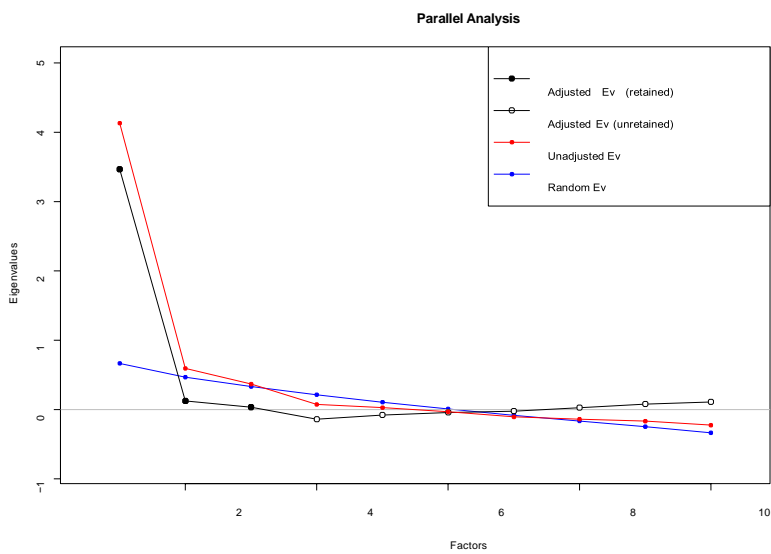
**Figure 2***Order effect of case presentations*

*Note.* Error bar = standard error. In condition A, raters first rated the authentic SP, followed by the unauthentic SP. In condition B, the opposite order was implemented.

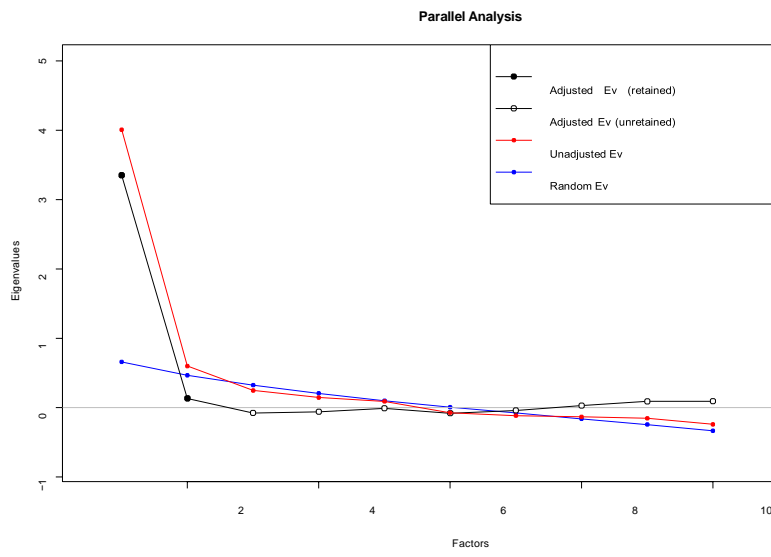
## Supplemental material



*Note.* Parallel analysis over both SPs ( $N = 97$ ). Ev = Eigenvalue; Adjusted eigenvalues = estimated eigenvalues adjusted for a finite sample size; Unadjusted eigenvalues = eigenvalues of the observed data from unrotated principal component analysis; Random eigenvalues = estimated eigenvalues from iterations number of random data sets.



*Note.* Parallel analysis for authentic SP ( $N = 97$ ); Ev = Eigenvalue.



Note. Parallel analysis for unauthentic SP ( $N = 97$ ); Ev = Eigenvalue.



## 7.2 Paper II

### **Can students in simulation portray a psychotherapy patient authentically with a detailed role-script? Results of a randomized-controlled study**

Destina Sevde Ay-Bryson<sup>a</sup>, Florian Weck and Franziska Kühne

Clinical Psychology and Psychotherapy, University of Potsdam, Germany

<sup>a</sup>Correspondence concerning this article should be addressed to Destina Ay-Bryson, Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany, Email address: [destina.ay@uni-potsdam.de](mailto:destina.ay@uni-potsdam.de)

#### **Copyright statement**

This is a raw version of the manuscript that was accepted for publication in “Training and Education in Professional Psychology”:

Ay-Bryson, D. S., Weck, F., Kühne, F. (2021). Can students in simulation portray a psychotherapy patient authentically with a detailed role-script? Results of a randomized-controlled study. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000388>

**Acknowledgments**

The authors would like to express their gratitude to Arne Klopffleisch for supporting the video laboratory, in which the present study took place. Further, we thank our interns Sophie Marie Schäfer and Juliane Pantke for their vital contribution during the video-ratings, as well as Samuel Oliver Ay-Bryson for programming support. Finally, we thank Marie Zorn (licensed psychotherapist) for her ratings of the SPs.

**Disclosure**

On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Funding**

The Hans Böckler Foundation supported this research with a full PhD scholarship to Destina S. Ay-Bryson.

### Abstract

**Objective.** The use of simulated patients (SPs) is increasing in popularity in psychotherapy training, but costs can limit their further spread. With a view to potentially lowering costs of recruitment and SP training, we examined whether a detailed patient role-script helped students to simulate a depressive patient more authentically than when they were only given basic information. **Methods.** A randomized-controlled study design was applied. Student SPs ( $N = 59$ ) were randomly allocated to an experimental (i.e., in-depth role-script) or control condition (i.e., text task). SPs' authenticity was evaluated from four different perspectives, including the SPs themselves, the study therapist involved in the simulations, two independent raters and a licensed therapist. **Results.** We found in-depth role-scripts to have a significant effect on SPs' authenticity ( $p < .001$ ). No correlation was found between SPs' authenticity with SPs' acting and psychotherapy pre-experience, empathy and personality. Authenticity correlated with extraversion ( $r = .36$ ,  $p < .01$ ) and the empathic concern scale ( $r = .27$ ,  $p < .05$ ) from self-report measures. **Conclusions.** The students were capable of authentically portraying a patient after receiving an in-depth role-script. Engaging students as SPs is thus reasonable. Detailed role-scripts could be easily implemented into curriculum for peer-simulations. Given the promising findings, future studies should examine the comparison between student SPs and real patients.

**Keywords:** Clinical psychology; education; evidence-based training; role-play; standardized patients

### Significance statement

This study demonstrates that student simulated patients (SPs) are capable of authentically portraying a depressive patient. Using detailed role-scripts and engaging students in simulation-based education could lower SP-associated costs.

## Introduction

Simulation-based education refers to learning, teaching and assessing environments and is understood as a technique to replace or complement experiences with real patients (e.g., Gaba, 2007). Often, simulation-based education includes a trainee and a simulated patient. The latter can be referred to with several terms, such as role-player, patient instructor, actor or standardized patient (Nestel & Bearman, 2015). “Simulated patients” are referred to as trained laypersons who may present symptoms within simulations (Beigzadeh et al., 2016; Kühne et al., 2018). Further, if simulated patients simulate symptoms repeatedly and in a standardized manner “with consistent, unvarying responses during the interaction” (Kühne et al., p. 773), they are often referred to as “standardized patients” (Adamo, 2003; Barrows, 1993; Pheister et al., 2016). There are also other approaches to defining standardized patients in contrast to simulated patients. For instance, Beigzadeh et al. (2016) defined standardized patients as actual patients or laypersons who are briefed to play a patient role using their own personal problem and presenting real feelings (e.g., emotional as well as personal characteristics). While it seems that there is some disagreement as to whether or not simulated or standardized patients should use own illnesses for portrayal (Kühne et al., 2018), the terms standardized and simulated patients are often used interchangeably. In the present paper, we focus on simulation-based education involving simulated patients (from now on referred to as “SPs”). Compared to real patients, SPs possess numerous benefits, such as repeatability and the opportunity to acquire skills in an environment that trainees may experience with increasing confidence (Pheister et al., 2016). While simulation-based education originated from medical education (Barrows & Abrahamson, 1964), it is widespread in a range of health profession fields, such as nursing education, pharmacy,

or physiotherapy (Nestel et al., 2015), and is now also disseminating more into clinical psychology and psychotherapy (Kühne et al., 2018).

Simulation-based education is considered a highly effective method of improving skills in health professional students (Cook et al., 2011, 2012; McGaghie et al., 2011), and is even associated with improved patient outcomes (Zendejas et al., 2013). Further, simulation-based education has been demonstrated effective specifically in mental health education (Piot et al., 2020; Rønning & Bjørkly, 2019). With regard to psychotherapy training, one study (Partschefeld et al., 2013) investigated whether the integration of SPs increases therapeutic skills, such as empathy and self-efficacy. In their study on a pre-post comparison, psychotherapy trainees' therapeutic skills were assessed from three perspectives, namely from the SPs', trainees' and from one external evaluator's. Results suggested a significant increase in therapeutic skills across all perspectives.

With a view to an increasing popularity of simulation-based education in psychotherapy training, adjustments are necessary. In a recent review, barriers and facilitators based on 41 studies were identified that should be considered while using SPs in the context of psychotherapy research and training (Kühne et al., 2018). Notably, the authenticity and realness of simulations was found to be of major concern. For instance, one pilot study (Krahn et al., 2001) found that less authenticity of SPs was associated with less empathy in the students. Authenticity can be defined as the "impossibility of distinguishing SPs from (real) patients" (Wünderlich et al., 2012, p. 501). Moreover, simulation-based education is associated with high costs (Kühne et al., 2018; Lane & Rollnick, 2007). High costs may stem from specialized faculty (i.e., experienced SP trainers and expert supervisors from the respective discipline; Hodges et al., 2002), required technology and, if applied, hiring professional actors. Hence, there is a call for

“cheaper alternatives that may be equally effective in facilitating the acquisition of communication skills” (Lane & Rollnick, 2007, p. 13).

Similarly, as there are various terms to the concept of simulation-based education and SPs, there also is a noticeable variety on who to engage as SP. While the advantages and disadvantages of SPs versus real patients appear more obvious (e.g., Bokken et al., 2010), the shortcomings and benefits of different SPs (e.g., faculty, students versus actors) appear less clear. However, it appears to be a more cost-effective approach to engage enrolled students as SPs compared to actors. A recent observational study (Pritchard et al., 2020) suggested that (physiotherapy) students were capable of realistically portraying a patient, promoting the further study of student SPs.

Commonly, thorough training of SPs alongside their proper recruitment is considered necessary for successful simulation-based education (Adamo, 2003). Furthermore, trainers of SPs should be experienced with SPs (Hodges et al., 2002). These factors arguably contribute to high costs in simulation-based education. However, the literature on SPs exhibits a vast heterogeneity with respect to SP training and, more specifically, role-scripts and instructions. Role-scripts can be defined as follows: “A detailed outline of the features to be portrayed by a standardized (or simulated) patient and detailed information about the life of the patient he or she is portraying.” (Hodges et al., 2002, p. 138). Nevertheless, there is no gold standard yet as to how to train SPs, nor a consensus on the role-script or instruction for SPs. For instance, some authors advocate role-scripts rich in details, whereas others argue for little information in the script, or even for instructing SPs to use a personal problem instead of using a script (Kühne et al., 2018; Klamen & Yudkowsky, 2002). Some studies used structured scripts for the training of specific psychotherapeutic skills (Coyle et al., 1998), and others used no specific script,

but clinical case vignettes (Imel et al., 2014). A recent review (Davies et al., 2020) confirmed that instructions used for SPs varied in terms of structure, length and depth.

Altogether, there is much variance in the way role-scripts are used in mental health simulation-based education, and no study to date examined the effects of different forms of role-scripts on the authentic portrayal of mental health problems when students are engaged as SPs. Thus, we intended to study whether student SPs would benefit from a detailed role-script on a patient case in contrast to little basic information in terms. Studying this would provide data on the potential merit of student-based simulation.

In this context, we intended to further explore potential associations between students' personality traits and pre-knowledge on mental disorders with their realistic portrayals as SPs. The scale perspective-taking of the Interpersonal Reactivity Index (IRI; Davis, 1980), which assesses facets of empathy, is defined as the spontaneous ability to adopt the perspective of others, while the scale fantasy is described as the tendency to transpose "imaginatively into the feelings and actions of fictitious characters in books, movies, and plays" (p. 114; Davis, 1983). Furthermore, openness of the Big Five Model (Goldberg, 1993; McCrae & John, 1992) has been found to be correlated with creativity (McCrae, 1987) and extraversion (Aluja et al., 2003), which may in turn be positively linked to authenticity.

Altogether, we aimed to test the following main hypothesis: We expected student SPs who received an in-depth role-script to be rated more authentic than when given only basic information, and also more authentic compared to the student SPs allocated to the control condition, who received only basic information. Further, we intended to explore the relation between authenticity and the personality of the students who were engaged as SPs, the students' empathy as well as their prior experience with mental health problems or psychotherapy, and with role-playing or acting.

## Methods

### Study design

To test our hypothesis, we used a randomized-controlled study design (see Figure 1), in which participants, i.e., bachelor and master students, portrayed a depressive patient with a standardized study therapist. All sessions were video recorded. Participants eligible to participate were randomly allocated to either the experimental ( $n = 29$ ) or control group ( $n = 30$ ) using random numbers generated on R version 3.4.2 (2017).

### Participants

#### *Simulated patients*

Enrolled students (engaged as SPs) fluent in German were eligible to participate. They were recruited via a participant platform of the University of Potsdam and received monetary reimbursement (10€) or course credit for participation. In order to prevent potential emotional distress due to portraying a patient, students with current mental health problems were excluded. All SPs provided informed consent to participate. One participant dropped out after receiving the baseline instruction. Hence, the final sample was  $N = 59$  ( $n = 43$ ; 72.88% female). Mean age was 24.34 years ( $SD = 4.12$ ;  $range = 17 - 35$ ); two of the participants were 17 years old; for these we also received informed consent by their parents. Half of the participants were Psychology students ( $n = 26$ ; 49.15%), the remaining SPs were students of different courses, e.g., sociology, mathematics or bioscience.

42.37% ( $n = 25$ ) reported to have had experience with acting, 20.34% ( $n = 12$ ) had experience with role-playing, which was mostly gained during school or lecture courses or in the context of university workshops. Finally, 42.37% ( $n = 25$ ) reported to have had contact with patients suffering from a mental health problem, and 45.76% ( $n = 27$ ) reported to have experienced psychotherapy, whereby (according to our ethics approval)



we did not specify whether experience occurred personally or with an acquaintance, who received treatment.

### **Procedure**

After filling in self-report measures, the participants were instructed to simulate a patient diagnosed with depression in a simulated therapy segment, lasting for approximately ten minutes. Participants were instructed to prepare themselves for the first simulation (pre) based on a basic role-script they were given (baseline), see Supplement 1 for details. After the first simulation the SPs received a sealed envelope with a task dependent on the group (experimental vs. control group) they were allocated to. SPs were given 15 minutes to either read through the in-depth role-script (experimental group) or the unrelated text (control group), see below for further details. Subsequently, the second simulation (post) followed. After the experiment was terminated, the SPs were provided with a task to help them step out of their roles in order to prevent potential induction of stressful emotions.

To enable comparability between the simulations, the study therapist (XX, psychotherapy trainee) was held standardized, and was blind to the status of the participants, which was ensured with the aforementioned sealed envelopes. The study therapist followed a previously developed script on a therapy session. The script covered the following: greeting the patient, setting agenda, exploring what the patient normally does in a day and whether they used to be more active in the past, and exploring the relation between activity and mood, and cognitions. The local ethics commission approved the study (no. 9/2018, University XX).

### ***Preparation for the simulations***

*Simulation pre.* For the baseline simulation (pre), all SPs received a basic script, which contained information on a fictional depressive patient with respect to diagnosis, therapy

context and duration, name, gender, age of the patient, appearance, current social and occupational situation, emotional situation and brief description of the therapy situation (see Supplement 1).

*Simulation post.* SPs of the experimental group received a written in-depth role-script with respect to more information in addition to the baseline information (see Supplement 1 for details). SPs allocated to the control group received an unrelated text task, which was a text about machine learning formulated for laypersons. Both texts were comparable in terms of length. To ensure participants had read the texts, all SPs were asked to answer three questions regarding the script or text they just read. The in-depth role-script was drafted by a licensed therapist (XX).

### **Raters and rater training**

#### ***Raters***

The portrayal of the participants was evaluated from four perspectives: first, the study therapist rated the SPs after each simulation. Second, two independent master's degree students in Clinical Psychology and Psychotherapy, i.e., student raters, rated the SPs after data collection was terminated; third, an independent licensed therapist (XX) from our outpatient clinic evaluated the SPs based on the post simulations. Fourth, all SPs were asked to self-evaluate their authenticity at the end of the experiment.

#### ***Rater training***

Prior to the external ratings all independent raters (i.e., student raters and licensed therapist) received a rater training. For the students this lasted four hours, while the licensed therapist accomplished the same training in 1.5 hours. The training consisted of two parts that are considered effective training components to prepare raters to accurately assess a certain behavior (Feldman et al., 2012). *Part I* included a performance dimension training during which raters familiarized with the targeted dimension, i.e., authenticity.

During this part of the training raters also received information on depression, including a video example, followed by a panel discussion. *Part II* consisted of a frame-of-reference training, during which four videos of role-play simulations were selected from a former pilot study and rated. Discrepancies between the raters were discussed. Then, the first five ratings of the actual ratings were compared between the raters, whereby these were not altered afterwards. The order of the videos (pre and post simulation) was randomized prior to the ratings, and to further minimize expectancy bias, the raters were blind to the temporal order of sessions as well as to the study design.

Agreement between both student raters for simulations at baseline was good ( $ICC_{(2,2)} = .84, p < .001, 95\% \text{ CI } [.75, .89]$ ), and also good for post ( $ICC_{(2,2)} = .87, p < .001, 95\% \text{ CI } [.81, .92]$ ) which suggests that the rater training sufficed. Agreement between the student raters and the study therapist was also good for both time points (pre:  $ICC_{(2,2)} = .88, p < .001, 95\% \text{ CI } [.81, .92]$ ; post ( $ICC_{(2,2)} = .89, p < .001, 95\% \text{ CI } [.83, .93]$ ). Further, the agreement between the student raters and the licensed therapist was moderate ( $ICC_{(2,2)} = .56, p < .001, 95\% \text{ CI } [.32, .72]$ ). Finally, agreement between the licensed therapist and the study therapist was moderate ( $ICC_{(2,2)} = .60, p < .001, 95\% \text{ CI } [.38, .74]$ ).

## Measures

An overview on which measures were filled or administered by whom is given in Supplement 2.

### *Big five inventory (BFI-10)*

In order to measure the personality of the students who portrayed a patient, the student SPs filled in the ten-item short version of the Big Five Inventory (Rammstedt & John, 2007). It covers all five facets: extraversion, agreeableness, conscientiousness, neuroticism, openness, and each item can be scored from 1 = disagree strongly to 5 =

agree strongly. In the current sample, Cronbach's  $\alpha$  ranged from .29 (agreeableness) to .76 (extraversion).

### ***Interpersonal reactivity index (IRI)***

To assess empathy, we further asked the SPs to fill in the IRI (original version: Davis, 1980, 1983, 1996; German version: Neumann et al., 2012). It contains 28 items and covers the following subscales: perspective-taking (i.e., ability to see a situation from someone else's perspective), fantasy (i.e., ability to identify with characters from books, movies or plays), empathic concern (i.e., ability to care about others' feelings and needs) and personal distress (i.e., ability to feel distress in difficult situations). A five-point scale is used ranging from 1 = does not describe me at all to 5 = describes me very well. In the current study, Cronbach's  $\alpha$  ranged from .61 (subscale empathic concern) to .79 (subscale perspective-taking).

### ***Experience***

Finally, the SPs answered four questions regarding their experience with acting, roleplaying, having suffered from mental health problems, and experience with psychotherapy. These items used a dichotomous (yes, no) answer option.

### ***Authenticity of patient demonstrations scale (APD)***

To assess the dependent variable we used the APD (XX et al., 2020), which is a rating scale, consisting of ten items. An item example goes as: "The person describes disorder-specific symptoms appropriately." It is scored on a 4-point Likert scale ranging from 0 = strongly disagree to 3 = strongly agree. Cronbach's  $\alpha$  ranged from .75 (self-report) to .95 (study therapist evaluator) for all different perspectives.

### ***Study therapist behavior check***

Since the same study therapist performed all simulations (59\*2 simulations) to keep them standardized, we additionally checked if the study therapist conducted the sessions

consistently. For this purpose, we randomly selected  $n = 30$  videos of the 118 simulations and instructed the two student raters to judge the study therapist with regard to the following items: (1) “The study therapist appears attentive”, and (2) “The study therapist follows the treatment script”. The items were to be scored on a 4-point Likert scale ranging from 0 = strongly disagree to 3 = strongly agree, whereby anchors for the scores 0 and 3 were provided.

### **Statistical analyses**

We conducted a priori power analyses using G\*Power (Faul et al., 2009). Since the study was conducted in a standardized laboratory setting with high internal validity, we expected a medium effect size. To detect a medium effect (Cohen’s  $f = .25$ ) for the interaction between time (pre and post) and condition (in-depth vs. basic role-script condition; repeated measures ANOVA, within-between-interaction) at a power of  $1 - \beta = .95$  with an  $\alpha$  of .05, a minimum sample of 54 participants was required. Considering potential dropouts, we aimed for a total sample of  $N = 60$ .

Prior to statistical analyses, data were checked for outliers; no participant was excluded. Due to one drop-out as mentioned above a final sample size of  $N = 59$  was considered in the following analyses. For measures with two raters (i.e., APD and study therapist behavior), we computed the mean over both raters. Group differences were examined with two sample  $t$ -tests. In order to evaluate the inter-rater reliabilities of the mean ratings of the APD, we computed intra-class correlation coefficients, 2-way random effects model  $ICC_{(2,2)}$  (Koo & Li, 2016; Shrout & Fleiss, 1979). According to Koo & Li (2016),  $ICCs$  between .50 and .75 indicate “moderate” and values between .75 and .90 indicate “good” reliability. To determine the interaction between the between-subject factor (i.e., condition) and the within-subject factor (i.e., time), we computed repeated-

measure ANOVAs. To explore the correlations between authenticity and the BFI-10 and the IRI, we computed Pearson' correlations. To explore the relation between authenticity and experience we calculated point-biserial correlations due to the variables for experience being dichotomous (Field et al., 2012). All statistical analyses were performed with R version 3.4.2 (2017); the alpha level was set to .05.

## Results

### Descriptive results and group differences

Altogether, 29 SPs were randomly allocated to the experimental group, and 30 to the control group. Mean age of the SPs allocated to the experimental group ( $M = 24.38$ ,  $SD = 4.70$ ) did not differ significantly from the mean age of SPs allocated to the control group ( $M = 24.30$ ,  $SD = 3.55$ ;  $t(52.12) = .07$ ,  $p = .94$ ).

Table 1 shows all APD mean scores and standard deviations. APD mean at baseline did not differ between both conditions (student evaluations:  $M_{\text{experimental}} = 1.32$ ,  $SD_{\text{experimental}} = .69$ ;  $M_{\text{control}} = 1.38$ ,  $SD_{\text{control}} = .68$ ;  $t(56.89) = -.34$ ,  $p = .73$ ; study therapist evaluations:  $M_{\text{experimental}} = 1.44$ ,  $SD_{\text{experimental}} = .80$ ;  $M_{\text{control}} = 1.39$ ,  $SD_{\text{control}} = .75$ ;  $t(52.51) = .23$ ,  $p = .82$ ). Nor did SPs' self-reported APD scores after the second simulation differ between both conditions ( $M_{\text{experimental}} = 2.18$ ,  $SD_{\text{experimental}} = .35$ ;  $M_{\text{control}} = 2.16$ ,  $SD_{\text{control}} = .30$ ;  $t(55.1) = .23$ ,  $p = .82$ ).

### In-depth role-script's effect on authenticity

#### *Student rater perspective*

From the student rater perspective there was a significant main effect of time  $F(1, 114) = 20.64$ ,  $p < .001$ ,  $\eta^2 = 1.3$ , as well as a main effect of condition  $F(1, 114) = 4.25$ ,  $p < .01$ ,  $\eta^2 = .06$ . Further, the repeated-measure ANOVA showed a significant interaction effect of Time\*Condition  $F(1, 114) = 13.66$ ,  $p < .001$ ,  $\eta^2 = .09$ ; see Figure 2 (Graph A). This model implies that there was a significant increase of authenticity at the second time point

in the experimental group in comparison to the control group, namely the group receiving an in-depth role-script.

### ***Study therapist perspective***

The second model we ran with APD study therapist measures as dependent variable resulted in a significant main effect of time  $F(1, 106) = 17.43, p < .001, \eta^2 = .12$  as well as a main effect of condition  $F(1, 106) = 14.57, p < .001, \eta^2 = .10$ . Further, a significant interaction effect of Time\*Condition was found  $F(1, 106) = 11.94, p < .001, \eta^2 = .08$ ; see Figure 2 (Graph B). This model implies that there was a significant increase of authenticity at the second time point, and that this increase differed significantly between both conditions, namely, that the group receiving an in-depth role-script was evaluated as significantly more authentic compared to the control condition.

### **Relation between authenticity and empathy, personality and pre-experience**

Altogether, the analyses showed that there is no correlation between authenticity (as measured by the student raters, the study therapist as well as the licensed therapist) and SP personality, empathy and pre-experience, see Supplement 3. However, authenticity as evaluated by the SPs themselves demonstrated a moderate correlation with self-rated extraversion ( $r = .36, p < .01$ ), implying that extraversion led to a higher evaluation of authenticity. The only facet of empathy that showed a significant relation with authenticity as evaluated by SPs was the empathic concern scale ( $r = .27, p < .05$ ).

### **Study therapist behavior check**

The mean of the item whether the study therapist was attentive was  $M = 2.93$  ( $SD = .12$ ) and mean of the item script adherence was  $M = 2.82$  ( $SD = .43$ ). The results indicate that the study therapist was comparatively attentive and adherent in all simulations.

## Discussion

In the current study, our main aim to study whether a detailed role-script would be more helpful for students to portray a patient authentically compared to basic information. In addition, we explored whether SPs' authenticity was related to their personality, empathy and prior experience (i.e., with acting and with mental health problems).

As hypothesized, SPs receiving an in-depth role-script were perceived as significantly more authentic compared to SPs who only received basic information. Compared to the baseline instruction, the in-depth role-script was more detailed and contained more information on specific aspects of the depressive patient case, such as concerning body language and reaction towards interventions. Possibly, student participants with arguably little patient experience needed this detailed information which enabled them to more realistically portray a patient. For instance, after receiving the in-depth role-script, it was most noticeable that SPs implemented a different posture and spoke more slowly and quietly compared to the first simulation. Further, during the second simulation, these SPs were more hesitant in trying out positive activities and their daily routines were described as more monotonous and less active. Consequently, it can be argued that if SPs implement the in-depth role-script correctly, they will achieve more realistic performances.

Our results are in line with previous studies that also engaged students as SPs. Partschefeld (2013) evaluated four psychology students as SPs, and in a more recent study, physiotherapy students ( $N = 40$ ) were taught to portray realistic patient roles (Pritchard et al., 2020). Both of these studies, however, included a minimum of 4 hours of SP training. Thus, given our comparable results regarding student portrayal, a detailed role-script may already be a good approach, albeit depending on the task. This in turn would imply that thorough planning and validating of role-scripts is crucially important.



It is desirable to replicate this finding for different therapeutic interventions and mental health problems.

Additional analyses showed no relation between authenticity and personality, empathy and prior experience with acting or role-playing or psychotherapy, with the exceptions of a relation between self-assessed authenticity and self-assessed extraversion, as well as the self-assessed empathic concern scale. One implication of these results may be that persons without prior experience may also be eligible to be an SP. Nevertheless, for future research it may be worthwhile to consider using full-length personality measures with psychometrically more sound properties (e.g., the internal consistency for agreeableness in the present sample was very low). Furthermore, the relation between SPs' prior experience with mental health problems and their authenticity could in future studies be investigated by comparing student SPs not only with one another, but also with trained laypersons or more experienced clinicians portraying a patient with a mental health problem.

Moreover, the agreement of APD ratings between the student raters and the study therapist were good. In contrast, the agreement between the student raters and the licensed therapist, as well as between the licensed therapist and the study therapist were only moderate. In the present study, the study therapist also trained the independent student raters for the authenticity evaluations. This may possibly account for the agreements. Therefore, future studies should particularly consider independent training situations for the raters. Still, student raters may supplement cost-intensive ratings by multiple external licensed therapists, as is often the case in different research areas of psychotherapy training (e.g., Kühne et al., 2019).

**Limitations and strengths of the study**

Limitations of the current study include that we only focused on depression as a mental health problem. To draw conclusions on SPs' authenticity more widely, investigations on more diagnoses would be necessary. Especially, it would be desirable to investigate portrayals of more complex mental disorders, such as psychoses or mania as these are considered to be more difficult to be portrayed by less experienced SPs (Kühne et al., 2018). Similarly, portrayals of comorbid patients should be considered as this is more prevalent in clinical practice in contrast to patients with only one mental disorder. Further, since one study therapist conducted all simulations to enable comparability, we are limited in investigating interaction effects with study therapist variables. SPs' authenticity may, however, depend on therapist behaviors, and vice versa, which should be investigated in future research. Moreover, since this study directly examined one component of SP training, it would be interesting for future research to compare in-depth role-scripts with more elaborate training that may include SP peers (i.e., others who are engaged as SPs), rehearsal and feedback regarding their authenticity. We emphasize that the current results should not be understood as a complete replacement of SP training, but rather argue that detailed role-scripts should be established as a standard in SP training. Nevertheless, role-scripts may suffice under certain circumstances, such as undergraduate curriculums.

Despite the limitations discussed, the study possesses various strengths that are worth being mentioned. First, we specifically investigated role-scripts by using a randomized-controlled study design. Second, in order to boost comparability, we kept all procedures of the study standardized. Specifically, sessions took place in a laboratory designed as a therapy room. Further, the study therapist followed a therapy script to also keep the

simulations comparable, whereby a certain degree of flexibility due to the SPs' reactions was inevitable. Third, we administered independent ratings with the student raters to check if the study therapist behaved comparably in the simulations. Fourth, for the student raters we conducted an elaborated rater-training, whereby we followed directions provided in the literature (see Feldman et al., 2012). Finally, agreements of evaluations were compared between four perspectives.

In conclusion, we demonstrated that students are capable of authentically portraying a patient when they are provided with a detailed role-script. Detailed role-scripts should therefore include information on, for instance, the reaction towards interventions, the body language, daily routine as well as the emotional and social situation. Considering the many tasks an SP is usually confronted within health education training, such as providing feedback and flexibly reacting towards trainees, additional SP training should be implemented, wherever possible.

### References

- Adamo, G. (2003). Simulated and standardized patients in OSCEs: Achievements and challenges 1992-2003. *Medical Teacher*, 25(3), 262–270. <https://doi.org/10.1080/0142159031000100300>
- Aluja, A., Garcia, O., & Garcia, L. F. (2003). Relationships among extraversion, openness to experience, and sensation seeking. *Personality and Individual Differences*, 35(3), 671–680.
- Ay-Bryson, D. S., Weck, F., & Kühne, F. (2020). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000349>
- Barrows, H. S. (1993). An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. *Academic Medicine*, 68(6), 443–451. <https://doi.org/10.1097/00001888-199306000-00002>
- Barrows, H. S., & Abrahamson, S. (1964). The programmed patient: A technique for appraising student performance in clinical neurology. *Journal for Medical Education*, 39(8), 802–805.
- Beigzadeh, A., Bahmanbijri, B., Sharifpoor, E., & Rahimi, M. (2016). Standardized patients versus simulated patients in medical education: Are they the same or different. *Journal of Emergency Practice and Trauma*, 2(1), 25–28. <https://doi.org/10.15171/jept.2015.05>.
- Bokken, L., Rethans, J.-J., Jöbsis, Q., Duvivier, R., Scherpbier, A., & van der Vleuten, C. (2010). Instructiveness of real patients and simulated patients in undergraduate medical education: A randomized experiment. *Academic Medicine: Journal of the*

*Association of American Medical Colleges*, 85(1), 148–154.  
<https://doi.org/10.1097/ACM.0b013e3181c48130>

Cook, D. A., Brydges, R., Hamstra, S. J., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hatala, R. (2012). Comparative effectiveness of technology-enhanced simulation versus other instructional methods: A systematic review and meta-analysis. *Simulation in Healthcare*, 7(5), 308–320.  
<https://doi.org/10.1097/SIH.0b013e3182614f95>

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *Jama*, 306(9), 978–988. <https://doi.org/10.1001/jama.2011.1234>

Coyle, B., Miller, M., & McGowen, K. R. (1998). Using standardized patients to teach and learn psychotherapy. *Academic Medicine*, 73(5), 591–592.  
<https://doi.org/10.1097/00001888-199805000-00058>

Davies, E. L., Pelentsov, L. J., Montagu, A., Gordon, A. L., Hooper, K. J., & Esterman, A. J. (2020). “Who Am I and Why Am I Here?” A Scoping Review Exploring the Templates and Protocols That Direct Actors in Their Roles as Simulated (Standardized) Patients. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*.

Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy*.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113.

- Davis, M. H. (1996). *Empathy: A social psychological approach, Vol.* Westview Press, Boulder, CO.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.
- Feldman, M., Lazzara, E. H., Vanderbilt, A. A., & DiazGranados, D. (2012). Rater training to support high-stakes simulation-based assessments. *Journal of Continuing Education in the Health Professions, 32*(4), 279–286. <https://doi.org/10.1002/chp.21156>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R.* Sage publications.
- Gaba, D. M. (2007). The future vision of simulation in healthcare. *Simulation in Healthcare, 2*(2), 126–135.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26.
- Hodges, B., Hanson, M., McNaughton, N., & Regehr, G. (2002). Creating, monitoring, and improving a psychiatry OSCE. *Academic Psychiatry, 26*(3), 134–161.
- Imel, Z. E., Baldwin, S. A., Baer, J. S., Hartzler, B., Dunn, C., Rosengren, D. B., & Atkins, D. C. (2014). Evaluating therapist adherence in motivational interviewing by comparing performance with standardized and real patients. *Journal of Consulting and Clinical Psychology, 82*(3), 472–481. <https://doi.org/10.1037/a0036158>
- Klamen, D. L., & Yudkowsky, R. (2002). Using standardized patients for formative feedback in an introduction to psychotherapy course. *Academic Psychiatry, 26*(3), 168–172. <https://doi.org/doi:10.1176/appi.ap.26.3.168>

- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.
- Krahn, L. E., Sutor, B., & Bostwick, J. M. (2001). Conveying emotional realism: A challenge to using standardized patients. *Academic Medicine*, *76*(3), 216–217.
- Kühne, F., Ay, D. S., Otterbeck, M., & Weck, F. (2018). Standardized patients in clinical psychology and psychotherapy: A scoping review on barriers and facilitators for implementation. *Academic Psychiatry*, *42*(6), 773–781.  
<https://doi.org/10.1007/s40596-018-0886-6>
- Kühne, F., Lacki, F. J., Muse, K., & Weck, F. (2019). Strengthening competence of therapists-in-training in the treatment of health anxiety (hypochondriasis): Validation of the Assessment of Core CBT Skills (ACCS). *Clinical Psychology & Psychotherapy*, *26*, 319–237.
- Lane, C., & Rollnick, S. (2007). The use of simulated patients and role-play in communication skills training: A review of the literature to August 2005. *Patient Education and Counseling*, *67*(1–2), 13–20.  
<https://doi.org/10.1016/j.pec.2007.02.011>
- McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, *52*(6), 1258.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215.
- McGaghie, W. C., Issenberg, S. B., Cohen, M. E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative

- review of the evidence. *Academic Medicine: Journal of the Association of American Medical Colleges*, 86(6), 706.  
<https://doi.org/10.1097/ACM.0b013e318217e119>
- Nestel, D., & Bearman, M. (2015). Introduction to simulated patient methodology. In *Simulated Patient Methodology: Theory, Evidence and Practice* (First Edition. Edited by D. Nestel and M. Bearman, pp. 1–4). John Wiley & Sons.
- Nestel, D., Morrison, T., & Pritchard, S. (2015). Scope of contemporary simulated patient methodology. In *Simulated Patient Methodology: Theory, Evidence and Practice* (First Edition. Edited by D. Nestel and M. Bearman, pp. 7–15). John Wiley & Sons.
- Neumann, M., Scheffer, C., Tauschel, D., Lutz, G., Wirtz, M., & Edelhäuser, F. (2012). Physician empathy: Definition, outcome-relevance and its measurement in patient care and medical education. *GMS Zeitschrift Für Medizinische Ausbildung*, 29(1).
- Partschefeld, E. (2013). *Evaluation des Einsatzes von Simulationspatienten in der psychotherapeutischen Ausbildung: Ein Beitrag zur empirischen Fundierung der Psychotherapieausbildung* [Doctoral dissertation].
- Partschefeld, E., Strauß, B., Geyer, M., & Philipp, S. (2013). Simulationspatienten in der Psychotherapieausbildung. *Psychotherapeut*, 58(5), 438–445.  
<https://doi.org/10.1007/s00278-013-1002-8>
- Pheister, M., Stagno, S., Cotes, R., Prabhakar, D., Mahr, F., Crowell, A., & Schwartz, A. (2016). Simulated Patients and Scenarios to Assess and Teach Psychiatry Residents. *Academic Psychiatry*, 114–117. <https://doi.org/10.1007/s40596-015-0477-8>
- Piot, M.-A., Dechartres, A., Attoe, C., Jollant, F., Lemogne, C., Layat Burn, C., Rethans, J.-J., Michelet, D., Cross, S., Billon, G., Guerrier, G., Tesniere, A., & Falissard,

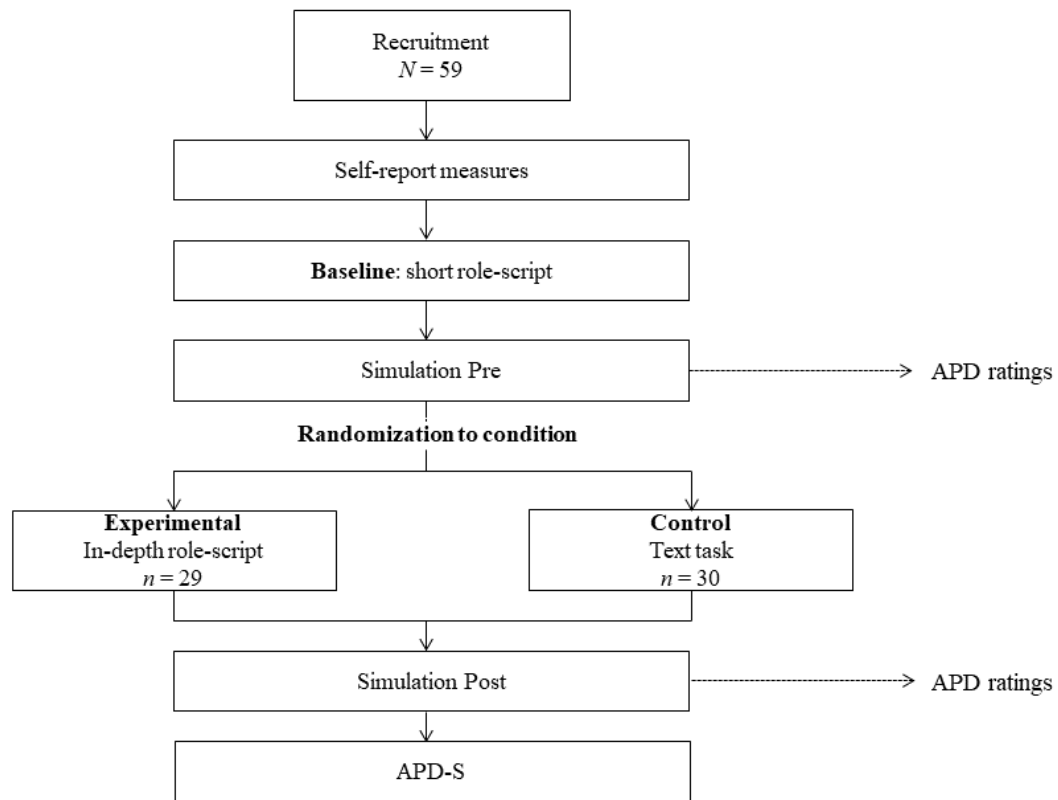


- B. (2020). Simulation in psychiatry for medical doctors: A systematic review and meta-analysis. *Medical Education*, 54(8), 696–708. <https://doi.org/10.1111/medu.14166>
- Pritchard, S. A., Keating, J. L., Nestel, D., & Blackstock, F. C. (2020). Physiotherapy students can be educated to portray realistic patient roles in simulation: A pragmatic observational study. *BMC Medical Education*, 20(1), 1–9.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212.
- Rønning, S. B., & Bjørkly, S. (2019). The use of clinical role-play and reflection in learning therapeutic communication skills in mental health education: An integrative review. *Advances in Medical Education and Practice*, 10, 415.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420.
- Wünderich, M. S., Nissen, C., Feige, B., Philipsen, A. S., & Voderholzer, U. (2012). Portrayal of psychiatric disorders: Are simulated patients authentic? *Academic Psychiatry*, 36(6), 501–502. <https://doi.org/10.1176/appi.ap.11090163>
- Zendejas, B., Brydges, R., Wang, A. T., & Cook, D. A. (2013). Patient outcomes in simulation-based medical education: A systematic review. *Journal of General Internal Medicine*, 28(8), 1078–1089. <https://doi.org/10.1007/s11606-012-2264->

**Table 1***Descriptive results of authenticity ratings and results of significance of group differences*

<u>Group</u>	<u>Evaluator perspective</u>					
	Student raters		Study therapist		Licensed therapist	Self-report
	Pre	Post	Pre	Post	Post	Post
	<i>M (SD)</i>					
Experimental	1.32 (.69)	2.31 (.53)***	1.44 (.80)	2.46 (.46)***	2.39 (.53)	2.18 (.35)
Control	1.38 (.68)	1.49 (.68)	1.39 (.75)	1.50 (.72)	2.02 (.60)	2.16 (.30)
Total	1.35 (.68)	1.90 (.73)***	1.42 (.77)	1.97 (.77)***	2.20 (.59)	2.17 (.32)

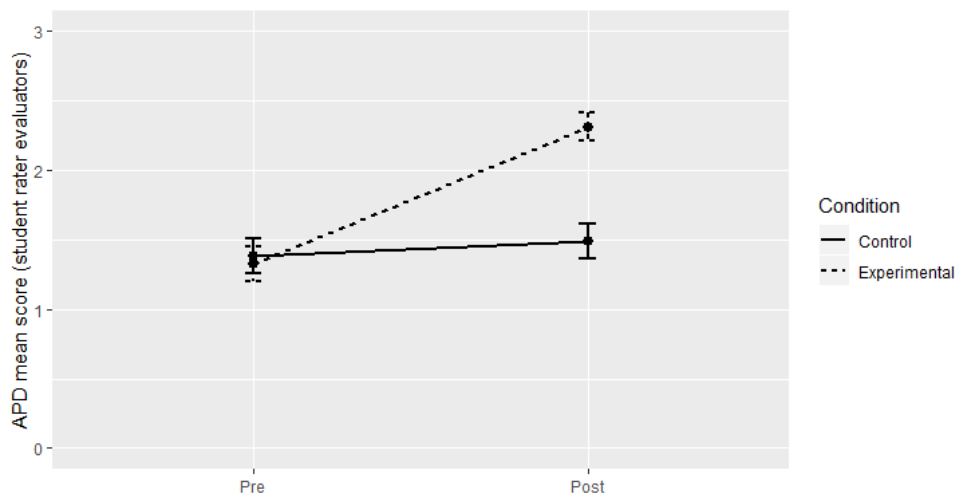
*Note.* A 4-point Likert scale ranging from 0 = strongly disagree to 3 = strongly agree was used; \*\*\* =  $p < .001$ .

**Figure 1***Study design*

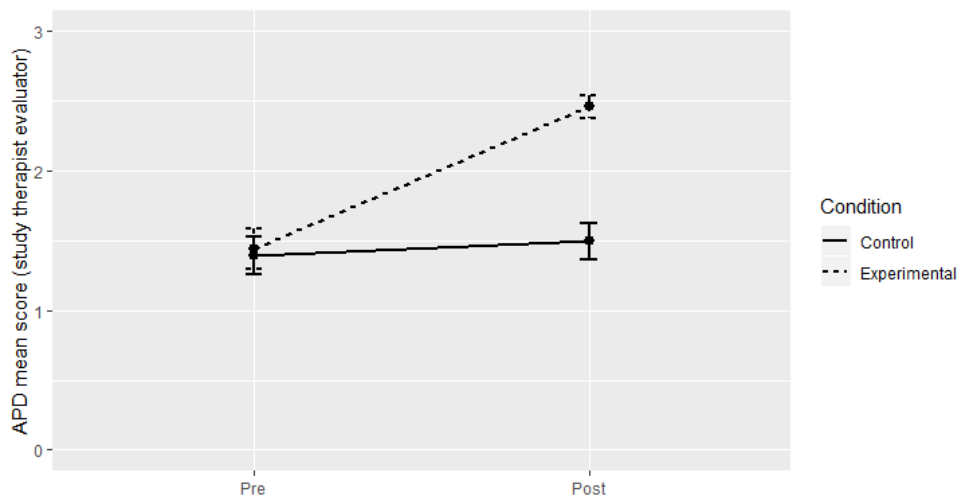
*Note.* APD = Authenticity of Patient Demonstrations, was applied by external raters based on video-recordings; APD-S = Self-version of the APD scale was only administered at simulation post in order to prevent bias regarding authenticity.

**Figure 2**

*Interaction effects of time and condition based on independent student rater (Graph A) and study therapist (Graph B) evaluations ( $p < .001$ )*



*Graph A*



*Graph B*

*Note.* Error bar = standard error.

**Supplement I Overview on script contents**

<b>Patient case</b>	<b>Domain</b>	<b>Content</b>
	Diagnosis	Depressive patient.
	Therapy context and duration	Therapy room, ca. 10 minutes.
	Name, gender, age	Ms./Mr. Miller, female/male, 25 years.
	Appearance	Reserved, young person, normally dressed.
<b>Baseline instruction</b>	Current social and occupational situation	Student, single, resident in studio, currently pausing side job at a café
	Emotional situation	For four months now you've been feeling utterly sad, useless, and powerless, in your eyes there's no explanation as to why you feel so depressed.
	Description of therapy situation	You've had appointments before, this session will be about how you spend the day and which activities may help you, the therapist begins the consultation.
	Personality	<b>all of the above and...</b> Usually you're an outgoing person, but currently it's difficult for you to stay in touch with friends. You avoid activities and dates, you feel sad and tired. You're conscientious and you feel embarrassed when you perceive results as flawed.
	Typical utterances that may be repeated	"I feel so powerless and tired." "I'm currently pausing my job at the café simply because the job is too exhausting for me at the moment."
	Reaction towards interventions	Initially you can't think of activities that used to be fun, after some hesitation you name cultural events and trips in the nature as activities you used to do. Currently you can only force yourself to do the most important things, such as grocery-shopping or going to doctor's appointments. Typically you reply "I tried that already, but I couldn't make it" when the therapists suggests activities.
<b>In-depth role-script</b>	Body language	Your head's hanging, and your posture is slumped, you avoid long eye contact with the therapist and your facial expression is sad. You speak slowly and quietly.
	Main complaint	You feel exhausted and depressed. Currently your job at the café is at pause because you don't have sufficient energy. You don't have much contact with friends. You only leave the flat when necessary. It all requires much more energy than before.
	Emotional situation	You suffer from losing contact with your friends and feeling uncomfortable being around other people. You feel guilty and feel like you strain others with your negative mood. When you meet up with friends you feel worthless.
	Social situation	You used to like your job at the café, which you also needed for financial reasons. Since you're close to the end of your studies you don't have to go to University frequently. Contact with friends decreased. In the past you enjoyed cultural events.
	Daily routine	You wake up quite early, but stay in bed until noon and ruminate. At daytime you watch series, this makes you forget your worries. You go to sleep quite early.

**Supplement 2***Overview on measures administered*

Measure	SPs	<u>Evaluator perspective</u>		
		Student raters	Study therapist	Licensed therapist
Sociodemographic data	x			
BFI-10	x			
IRI	x			
APD	x <sup>a</sup>	x	x	x
Study therapist behavior		x		

*Note.* x = administered by; BFI-10 = Big five inventory; IRI = Interpersonal reactivity index; APD = Authenticity of demonstrations scale; <sup>a</sup> = Adapted version to self-report; Study therapist behavior = evaluation of attentiveness and script-adherence of the therapist; SPs = Simulated patients.

**Supplement 3**

*Pearson's correlations between authenticity and personality, empathy and pre-experience (N = 59)*

APD evaluator	t	BFI-10					IRI				Experience with			
		Extra	Agree	Consc	Neuro	Open	Pers	Fant	Conc	Dist	Act	Rolep	Pat	Psychth
Student raters	Pre	-.25	-.16	.06	-.19	.17	-.15	.09	-.11	-.02	-.05	-.12	.19	-.02
	Post	-.08	-.14	.03	-.16	.04	-.17	-.01	-.00	-.06	-.14	-.13	.07	-.02
Study therapist <sup>1</sup>	Pre	-.20	.02	.01	-.19	.27	-.00	.05	-.15	-.04	.03	.04	.13	-.03
	Post	.03	-.09	.03	-.11	.08	-.00	.12	.03	-.12	-.03	-.02	.03	.04
Lic. therapist	Pre	.07	.02	-.14	-.14	.18	-.06	-.03	.03	.01	-.05	-.04	-.13	.15
	Post	<b>.36**</b>	.11	.22	-.05	.05	.24	.12	<b>.27*</b>	.00	-.04	.17	-.09	.06
SPs	Post													

*Note.* \* =  $p < .05$ , \*\* =  $p < .01$ ; <sup>1</sup> = Data of APD therapist ratings are missing for the first four participants; SPs = Simulated patients; t

= Time point; BFI-10 = 10-item short version of the Big Five Inventory; Extra = Extraversion; Agree = Agreeableness; Consc =

Conscientiousness; Neuro = Neuroticism; Open = Openness; IRI = Interpersonal Reactivity Index; Pers = Perspective-Taking Scale;

Fant = Fantasy Scale; Conc = Empathic Concern Scale; Dist = Personal Distress Scale; Act = Experience with Acting; Rolep = Role-

playing, Pat = Mental Health Patients; Psychth = Psychotherapy.

### 7.3 Paper III

#### **Can psychotherapy trainees distinguish standardized patients from real patients? Results of a pilot study**

Destina Sevde Ay-Bryson<sup>1,a</sup>, Florian Weck<sup>1</sup>, Peter Eric Heinze<sup>1</sup>, Thomas Lang<sup>2,3</sup> and Franziska Kühne<sup>1</sup>

<sup>1</sup>University of Potsdam, Clinical Psychology and Psychotherapy, Potsdam, Germany

<sup>2</sup>Jacobs University Bremen, Psychology and Methods, Bremen, Germany

<sup>3</sup>Christoph-Dornier-Stiftung, Bremen, Germany

<sup>a</sup>Correspondence concerning this article should be addressed to Destina Ay-Bryson, Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany, Email address: destina.ay@uni-potsdam.de

#### **Copyright statement**

This is a raw version of the manuscript that was accepted for publication in “Zeitschrift für Klinische Psychologie und Psychotherapie”:

Ay-Bryson, D. S., Weck, F., Heinze, P. E., Lang, T., Kühne, F. (in press). Can psychotherapy trainees distinguish standardized patients from real patients? Results of a pilot study. *Zeitschrift für Klinische Psychologie und Psychotherapie*. <https://doi.org/10.1026/1616-3443/a000594>



**Acknowledgments**

We would like to thank all psychotherapy trainees, SPs and patients for their participation in the study as well as Samuel Oliver Ay-Bryson for proofreading the manuscript.

**Disclosure**

The authors have no conflicts to disclose.

**Funding**

The Hans Böckler Foundation supported this research with a full PhD scholarship to Destina Sevde Ay-Bryson.

### Abstract

**Background.** Standardized patients (SPs) will become a standard component in psychotherapy education according to the new psychotherapy law in Germany, but little is known about their authenticity. **Objective.** Goal of the present pilot study was to explore if SPs, following an exhaustive two-day SP training, can be distinguished from real patients by psychotherapy trainees. **Methods.** Twenty-eight psychotherapy trainees ( $M = 28.54$  years of age,  $SD = 3.19$ ) participated as blind raters. They evaluated six video-recorded therapy segments of trained SPs and real patients using the Authenticity of Patient Demonstrations scale. **Results.** Authenticity scores of real patients and SPs did not differ ( $p = .43$ ). Descriptive results indicated that the highest score of authenticity was given to an SP. The real patients did not differ significantly from the SPs with regard to perceived impairment ( $p = .33$ ) and the likelihood of being a real patient ( $p = .52$ ). **Conclusions.** The current results suggest that the SPs were not distinguished from real patients by psychotherapy trainees. We therefore strongly recommend incorporating training prior to utilization. Limitations and future research directions are discussed.

**Keywords:** Evidence-based training; learning; simulated patients; simulation-based education; therapist competence

### Introduction

According to the new psychotherapy law in Germany (Psychotherapeutenausbildung, 2019), standardized patients (SPs) will become a standard component in psychotherapy education similar to medical education, where SPs have been integrated for decades as part of the objective structured clinical examination (e.g. Adamo, 2003). From a theoretical point of view, there is ongoing discussion on the role of assessment and training of therapist competence in the field of clinical psychology. In a recent review, Muse and McManus (2013) extended Miller's (1990) hierarchical framework for clinical assessment with the measurement of cognitive behavioral therapist competence. According to the framework, different levels of clinical competence can be assessed in different ways. *Knowledge*, representing the first level on the hierarchy, may be assessed by essays or multiple choice questions (Muse & McManus, 2013). Assessing *practical understanding* represents the second level, which may additionally be evaluated by case reports or short-answer clinical vignettes. The penultimate level on the hierarchy aims to answer the question whether the therapist can *demonstrate the skill*, followed by the final level, *clinical practice*, referring to the therapist's use of the skill. The penultimate level becomes especially relevant for psychotherapy training. To assess practical application of knowledge, i.e., skill, Muse and McManus (2013) referred to the utilization of standardized role-plays, which they define as "artificial simulations of clinical scenarios in which a therapist interacts with an individual playing the role of a standardized patient" (p. 490).

SPs were first introduced by Barrows and Abrahamson (1964) to evaluate students' clinical performances in clinical neurology and they are widely accepted in the medical field (Kühne et al., 2018). SPs are healthy laypersons who portray a clinical problem in a standardized manner and, although they have been regarded with skepticism in the

context of psychotherapy training and research (Hodges et al., 2014; Kühne et al., 2020), they are currently disseminating more into clinical psychology and psychotherapy (Kühne et al., 2020). For instance, one previous study (Partschefeld et al., 2013) investigated the integration of SPs in psychotherapy training and found promising effects with regard to therapeutic skills. Further, following the new psychotherapy law, SPs will likely be integrated even more across Germany. While the use of SPs for the assessment of competence has promising potential for training and research, it has also been noted that it potentially simplifies clinical complexity which may hinder authenticity (Sharpless & Barber, 2009).

Due to increasing attention towards SPs in research and training (Melluish et al., 2007; Partschefeld et al., 2013; Sheen et al., 2020), it is crucial to examine authenticity of SPs in order to overcome reservations associated with simulations' lack of realness. We refer to the definition of authenticity proposed by Wünderich et al. (2012), namely the "impossibility of distinguishing SPs from (real) patients" (p. 501). Few studies have examined the authenticity of SPs by comparing them with real patients. In the often-cited pilot study by Krahn et al. (2002), SPs were provided with a case in an outline format, which enabled them to improvise their answers during simulated interactions. SPs received one training session before the actual simulation, however the duration of the training was not specified. Students who conducted interviews with patients could correctly identify the SPs most of the time. The results of this study also showed that 91% of students who believed the patient to be an SP felt less empathy towards the patient. The authors of this pilot study concluded that "training must focus on facilitating actors' ability to convey emotion realistically and therefore evoke empathy in the interviewer" (p. 30).

In contrast, in a more recent study (Wünderich et al., 2012) SPs were trained for about four hours in order to simulate a patient authentically. Here, experienced psychiatrists were asked to assign one of three labels (SP, patient, unsure) retrospectively to each interviewed person. Their results demonstrated that, although SPs were rated as less authentic compared to real patients, they were not detected in 70 of 114 (61.40%) SP cases. Wünderich and colleagues concluded that SPs “with proper training, can reach a high level of authenticity in presenting major psychiatric disorders when rated by experienced psychiatrists” (p. 501). Although it seems plausible that training SPs would result in more authentic portrayals of patient cases (Partschefeld, 2013), the literature provides little evidence for this assumption. Consequently, there is no gold standard yet as to how to train SPs.

Although both studies (Krahn et al., 2002; Wünderich et al., 2012) examined authenticity of SPs as evaluated by external judges, they both concentrate solely on subjective, single-item assessments (e.g., retrospective allocation to group) and medical related samples (e.g., psychiatrists). Finally, the simulated interactions were not standardized, which limits comparability.

### **Objective**

In the current pilot study, we aimed to address these limitations and developed a two-day SP training. In a previous randomized-controlled study (Ay-Bryson et al., under review), we found that SPs can be trained to be more authentic with a detailed role-script on a patient case compared to basic information. In the current pilot study, we aimed to explore whether SPs, following a thorough training, can be distinguished from real patients by psychotherapy trainees. We addressed the following research questions: (1) Can psychotherapy trainees distinguish between trained SPs and real patients? Further, since an SP can be evaluated as authentic, but still not be perceived as a real patient, we

were interested in the following research questions: (2) Is there a relationship between authenticity and the likelihood of the interviewed person being a real patient; and is there a relationship between authenticity and their psychological impairment?

## **Methods**

### **Study design**

Participants who were eligible to participate (i.e., raters; see below) watched six video-recorded simulations of five-minute therapy segments. The raters were instructed to evaluate the interviewed persons in the simulations with regard to their authenticity and psychological impairment. Participants were not informed about the status of the interviewed persons (real vs SPs), nor were they informed that the videos would include SPs and/or real patients, which was ensured by using the term “person” in the instruction.

### **Participants**

#### ***Raters***

A total of 29 raters participated in the current study, of whom one was excluded from analyses due to too many missing data (i.e., answered only half of the items). Hence, the final sample was  $N = 28$  (82.14% female). Persons currently undergoing psychotherapy training or licensed psychotherapists were eligible to participate; all participants provided informed consent prior to participation. The mean age of the raters was 28.54 years ( $SD = 3.19$ ;  $range = 24 - 40$ ). Most of them were psychotherapy trainees ( $n = 26$ ; 92.86%), all of whom were specialized in cognitive behavioral therapy. The average duration of psychotherapy training was 8.73 months ( $SD = 2.40$ ;  $range = 4$  to 10 months).

#### ***Standardized and real patients***

The mean age of the four female SPs was 22.75 years ( $SD = 3.63$ ;  $range = 20 - 29$ ) and they reported no prior experience as SPs before taking part in the current study. The SPs were part of an ongoing research project (Kühne et al., 2020). They were selected

based on the following criteria: no mental illness, ability to reflect, and joy doing theatrical work. The two real patients aged 23 (female) and 24 (male), both diagnosed with depression (recurrent and episode, respectively), were undergoing treatment at the outpatient unit of the Department of Clinical Psychology and Psychotherapy at the University of Potsdam. We recruited these patients in accordance with the following questions: Do they fulfil the criteria for mild depression? Are they available? Do they consent to being video-recorded? Does their sociodemographic data match that of the SPs? The real patients received background information on the study as well as a description of the situation, namely that the simulation would take place in the laboratory in order to ensure comparability of the simulated interactions. The therapy situation was a therapist exploring a specific situation typical for depression reported by the SP/real patient. In accordance, SPs portrayed a patient with a first depressive episode. All SPs and real patients provided informed consent prior to participation.

### **Procedure**

Our university's ethics committee approved the study (no. 9/2018, University of Potsdam).

### ***Setting***

The six video segments were recorded in the laboratory designed as a therapy room at the University of Potsdam, see Figure 1. In order to enable comparability, all sessions were conducted by one licensed psychotherapist (FK). Four of the sessions featured a trained SP and two featured real patients, whereby the sessions focused on the exploration of a specific situation. The ratings by psychotherapy trainees took place at the Institute for Psychological Psychotherapy Training in Bremen (IPP) and at the University of Potsdam by psychotherapy trainees of the Psychological-Psychotherapeutic Institute

(PPI) at the UP Transfer. All video segments were viewed sequentially. The order of the video segments and the duration of each video was held consistent for all raters.

### ***SP training***

With respect to training SPs to enable them to simulate their patient roles authentically (i.e., indistinguishably from real patients), SPs underwent a two-day workshop at the University of Potsdam, which altogether, including homework, took 12 hours. The training consisted of seven units: (1) project introduction, (2) SP concept, (3) psychoeducation, (4) role descriptions, (5) authenticity, (6) role analysis and comparability, and (7) role-play exercises, de-roling and feedback. We conceptualized the SP training following previously published manuals (Brem et al., 2018; Scherer & Ehrhardt, 2017) as well as published overviews (Adamo, 2003; Peters & Thrien, 2018; Voderholzer, 2007).

In *project introduction*, the accompanying research project (Kühne et al., 2020) was presented and all staff members were introduced. During *SP concept*, the definition of an SP as well as their benefits and drawbacks for psychotherapy training were discussed. Subsequently, participants were made familiar with their tasks, including the scripted scenarios of a depressed patient. *Psychoeducation* included general information about mental disorders followed by a brief introduction to cognitive behavioral therapy (CBT). Diagnostic criteria of depression were then presented and to further clarify the manifestation of depressive symptoms, a case example was discussed. Furthermore, participants were informed about treatment options for depression with a focus on CBT followed by information regarding particular therapeutic strategies, such as changing behavior and changing thoughts (Hautzinger, 2013). Six different *role descriptions* had to be read carefully at home by all participants, which were then discussed with regard to diagnostic criteria of the portrayed disorder and difficulties in the portrayal. Different



aspects of *authenticity* were discussed (“In your opinion, what makes a portrayal authentic?”), which aimed to clarify the individual concepts of authenticity. It was further explained that authentic SPs present cases more typically than real patients and that those representations are often more tangible for students or trainees. The discussion aimed to clarify advantages of authentic SP role-plays as facilitators of learning conditions. Furthermore, to create awareness of symptoms typical for depression, the specific diagnostic criteria were conveyed. In order to improve authenticity, a video analysis of a role model and of the SPs themselves was implemented, and positive and negative performances were discussed (“How convincing did you find the SP?”). During *role analysis and comparability*, different role descriptions were analyzed by discussing questions about demographic facts, symptoms, social environment, biography, thoughts and feelings of the scripted person. Participants were encouraged to think of strategies for carrying out an authentic role-play of each character. Subsequently, it was discussed that the SPs should make an effort to simulate the patient in a consistent and hence comparable manner during the subsequent interactions. Finally, *role-play exercises, de-roling and feedback* involved practicing basic acting skills. Acting exercises were conducted with a focus on nonverbal acting. This was followed by the introduction of de-roling techniques, such as sitting on a different chair, taking off clothes and requisites associated with the role, and discussing the latest session with a peer. Finally, role-plays were carried out and evaluated and feedback was given by a licensed psychotherapist (FK) to the SPs.

### ***Simulated situations and scenarios***

The simulated situations were based on scripted scenarios. For the development of the scripts, we followed those previously published (Voderholzer, 2007). The scripts included information regarding depressive symptoms, factors that contribute to the development and maintenance of depression as well biographical aspects (for a scenario

example, see Supplement 1; and Kühne et al., 2020). The main content of the scripts was paralleled. However, the background story of the SPs differed slightly in order to make SPs not too obvious compared to real patients. For example, one SP portrayed a person feeling overwhelmed with the question of what to do after school graduation, whereby another SP portrayed a person experiencing perinatal depression. The role-scripts were previously evaluated by an independent licensed psychotherapist and a real depressive patient (not included in the study) regarding comprehensibility, feasibility, relevance and suitability for an authentic portrayal of a depressive disorder.

## **Measures**

### ***Authenticity of patient demonstrations scale (APD)***

To assess the authenticity of SPs and real patients, we used the 10-item Authenticity of Patient Demonstrations scale (APD; Ay-Bryson et al., in press). One example item is: “The person describes disorder-specific symptoms appropriately”. It can be scored on a 4-point Likert scale ranging from 0 = “strongly disagree” to 3 = “strongly agree”. The APD was demonstrated in a previous study (Ay-Bryson et al., in press) to distinguish well between authentic and unauthentic SPs, to have a one-factorial structure, and was found to have good internal consistency (Cronbach’s  $\alpha = .83$ ). Further, evidence of convergent and discriminant validity exists (Ay-Bryson et al., in press), as it correlated ( $r = .82$ ) with an established tool for the assessment of SP performance in medical contexts and demonstrated no correlation with unrelated symptoms (i.e. anxiety) that were not supposed to be portrayed ( $r = .004$ ). In the current study Cronbach’s  $\alpha$  ranged from .84 to .95. The internal consistency is comparable to that of the original study of the scale (Ay-Bryson et al., in press) and can be considered good to excellent. Interrater reliability between all raters over the whole sample was  $ICC_{(2,28)} = .76, p < .001, 95\% CI [.49, .94]$

(Koo & Li, 2016). According to Koo & Li (2016), *ICCs* between .75 and .90 indicate good reliability.

### ***Additional items***

Further, we asked the raters to evaluate *how psychologically impaired* (i.e., impairment) each patient appeared. Finally, raters estimated *how likely* it was that each interviewed person was a *real patient* (i.e., real patient). Both items used an 11-point scale ranging from 0 to 100%.

### **Statistical analyses**

A final sample size of  $N = 6$  simulated interactions based on  $N = 28$  raters was included in statistical analyses. Two reversed-scored items of the APD were inverted. Then, we computed means over all raters for each SP/real patient, and calculated group means for the comparison of SPs vs. real patients.

In order to test whether SPs and real patients differed, we computed two sample *t*-tests. To explore the relationship between authenticity and the additional two items, we computed Pearson's correlations and used the *Bonferroni* corrected *p*-values (see Table 2). All statistical analyses were performed with R version 3.4.2 (2017); the alpha level was set to .05.

## **Results**

### **Indistinguishability of SPs (research question 1)**

Mean scores of the APD and additional items as well as grouped means are given in Table 1. On an average, the raters "agreed" that the SP group ( $M = 2.10$ ,  $SD = .35$ ) as well as the real patients group ( $M = 2.01$ ,  $SD = .49$ ) portrayed their roles authentically. The raters indicated a 51% ( $SD = .14$ ) probability for the SPs to be a real patient, and a 55% probability for the real patients. Finally, SPs were evaluated to have a 50% (.13)

psychological impairment, whereby the real patients were evaluated as slightly less impaired ( $M = 46\%$ ,  $SD = .16$ ).

### ***Authenticity***

APD means of SPs did not differ significantly from the APD means of the real patients;  $t(48.93) = -.79$ ,  $p = .43$ ; Cohen's  $d = -.21$ , 95% CI  $[-.75, .33]$ .

### ***Realness and impairment***

SPs and real patients did not differ significantly on the item "real patient";  $t(43.79) = .66$ ,  $p = .52$ ; Cohen's  $d = .18$ , 95% CI  $[-.36, .71]$ . With regard to the item "impairment" SPs and real patients did not differ significantly;  $t(51.92) = -.98$ ,  $p = .33$ ; Cohen's  $d = -.26$ , 95% CI  $[-.80, .28]$ .

### **Correlations between authenticity, realness and impairment (research question 2)**

Pearson's correlations between the APD and the additional items are presented in Table 2. As expected, authenticity strongly correlated with the item "real" ( $r = .62 - .76$ ,  $p < .001$ ). On the other hand, item "impairment" did not correlate significantly with authenticity ( $r_{\text{realpatient}} = .13$ ,  $p = .51$ ;  $r_{\text{SP}} = .19$ ,  $p = .35$ ). Further, there were strong, but non-significant correlations between both additional items ( $r_{\text{SP}} = .43$ ,  $p = .02$ ;  $r_{\text{realpatient}} = .52$ ,  $p = .004$ ).

## **Discussion**

With a view to facilitating an increasing spread of SPs worldwide but also particularly in Germany, in the current paper we suggest careful training of SPs prior to their deployment. The major aim of the current pilot study was to explore if psychotherapy trainees can distinguish between real patients and SPs who have been trained in an elaborate two-day training. Furthermore, in this study, we observed no significant difference between real patients and SPs regarding authenticity, perceived impairment and the perceived likelihood of being a real patient. Still, we caution against overreliance

on the results, and propose to use equivalence tests based on bigger sample sizes in order to determine the absence of an effect (Anderson & Maxwell, 2016; Lakens, 2017).

Pilot results of the current study suggest that psychotherapy trainees could not distinguish trained SPs from real patients. In fact, considering descriptive results, one of the two real patients was rated the least authentic of all interviewed persons, and the person with the highest authenticity score happened to be an SP. As a consequence of these results, we conclude that the training we conducted was effective in equipping SPs with the ability to portray a depressive patient in an authentic manner. This may be due to the theoretical input regarding background and conceptualization of SPs, and to the practical phase of the training, during which SPs received feedback by a licensed psychotherapist (FK). It may also be the case that SPs could better study their roles because they received an introduction into the theoretical background of the clinical picture of depression, which included general knowledge as well as specific diagnostic criteria and treatment options (Hautzinger, 2013). However, in order to draw firm conclusions on the SP training proposed in our study, we suggest future studies to conduct pre-post training comparisons of authenticity.

In line with our expectation, persons who were rated as authentic were also evaluated as likely to be a real patient. We interpret this result as further evidence that the raters could not tell SPs from real patients. Although the correlation between psychological impairment and the likelihood of being a real patient was not significant after a *Bonferroni* correction, this relation could be investigated in future studies based on bigger sample sizes. Arguably, perceived impairment plays a central role in the context of authenticity. Interestingly, Wünderich and colleagues (2012) found in their study that SPs were rated significantly better for the items “case is typical” and “symptoms are obvious to students”.

Since clinically significant impairment is one main criterion, it is conceivable that the introduction to specific diagnostic criteria is of major importance when training SPs.

Similar to previous studies (Krahn et al., 2002; Wünderich et al., 2012) we compared trained SPs with real patients, but adapted this approach to psychotherapy training. Notably, we ensured comparability between simulated interactions by standardizing the therapist who conducted all simulations. Finally, we considered psychotherapy trainees as raters, unlike in the study of Wünderich et al. (2012) in which experienced psychiatrists rated SPs. Arguably the perspective of trainees is relevant considering that it is mostly they who are to be trained and assessed in interactions with an SP. On the other hand, given that the training with SPs aims to prepare trainees for real clinical encounters, experienced clinicians should be consulted too in terms of evaluating the representativity and authenticity of SPs. Ideally, future studies should incorporate both perspectives (i.e., trainees and experienced clinicians).

Considering average descriptive results of the item real, we noticed that trainees were rather unsure whether the interviewed person was a real patient. This may potentially be associated with the early stage of training as the majority of the raters were in their first year of training. On the other hand, those who evaluated the interviewed person as authentic also evaluated the person to likely to be a real patient, which is also demonstrated in the strong correlations. Therefore, it seems crucial to focus on the aspect of authenticity during SP training. An alternative explanation for the uncertainty may also be that although we successfully trained SPs to be evaluated as authentic, there is still room for improvement of performances. Higher levels of authenticity may, for instance, be reached through refresher training sessions, through inviting real patients to a training session or by using model learning situations adjusted to the scenarios that are to be simulated. However, it could also be the case that an SP is authentic but not believed to

be a real patient. It would be interesting to investigate whether this would contribute differently to learning effects. Precisely, future studies could examine whether the perceived authenticity has a stronger effect on learning than perceived believability of the SP. It is conceivable that trainees would still benefit from training with authentic SPs even if they are not believed to be real. If this were the case empirically, one consequence would be that masking the status of the SP would not be necessary for training purposes.

### **Limitations and future research directions**

While the results of the current pilot study are promising, limitations should be considered. Importantly, future studies should conduct a priori power analyses in order to determine the necessary sample size. Since the current pilot study considered a small sample size, the generalizability of the herein reported results is limited. Consequently, future studies should ideally consider a bigger sample size of SPs and real patients. Another limitation of the present study is that the groups were unbalanced (i.e., 4:2 SPs:real patients). This was due to patient availability at our outpatient clinic. A strength of the current study was that we successfully matched the overall sociodemographic data of the real patients with that of our SPs and that the clinical picture was comparable, too. Despite our data's clustered structure (i.e. each psychotherapy trainee rated all patients), we could not conduct analyses that consider multilevel data due to our sample size (Maas & Hox, 2005). Consequently, future studies would benefit from considering bigger sample sizes also in this manner.

Further, we considered one clinical picture only, i.e., depression. Thus, the present results should be replicated on the basis of different mental health problems. Arguably, different mental illnesses may vary in the difficulty to portray them authentically. For instance, more complex mental disorders, such as mania or psychoses, seem more difficult to portray (Kühne et al., 2018). Similarly, it may be more challenging to portray

patients with comorbid disorders. Studying this, however, would be highly desirable as comorbidity is usually prevalent in clinical encounters (e.g., Andrews et al., 2001; Angst et al., 2004). It is conceivable that experienced SPs with longer histories of being engaged as an SP are more likely to portray complex disorders more authentically. Moreover, both additional items, i.e., real and impairment, are single-items and should therefore be interpreted with caution. Future studies should consider briefing raters beforehand, in order to prevent potential confusion regarding these variables.

For economic reasons, we included five-minute therapy segments on focused interactions. Thus, the results deliver first clues but should be expanded in future studies. In particular, it would be relevant to replicate the present results based on longer segments or even full sessions.

Future studies may also benefit from conducting dismantling studies to identify which of the training components accounted for authenticity (e.g., Ay-Bryson et al., under review). Such results may be favorable for SP programs with less time and fewer resources in order to adapt the length of the training to availability.

Finally, raters of the current study were blind to the status of the interviewed persons. Indeed, they were not informed that some of the patients would be SPs and some real patients, unlike in the study of Wünderich et al. (2012). It is possible that the instruction given to raters has an effect on how they perceive the interviewed persons. When conducting observation studies, we should consider potential biases, such as the halo or contrast effect, that may contribute to systematic errors in observation (Gräf & Unkelbach, 2016; Wirtz, 2017). Similarly, an expectancy effect may influence our evaluation of an observation. For instance, a former study found an expectancy effect with regard to the rating of severity of current symptoms portrayed by SPs (Mumma, 2002), whereby the ratings were influenced by prior information about the SPs. As a



consequence, it may be the case that the SPs of the current study would have been detected if the trainees had been given the information beforehand. On the other hand, in the study of Wünderich et al. (2012), it was noticed that “in the majority of the cases, SPs were not detected” (p. 501). Nevertheless, several experimental studies highlight the importance of controlling for and reducing rater expectancy and related effects (e.g., Ariel et al., 2019; Hoyt, 2002; Martell & Evans, 2005). Thus, it would be interesting and important for future studies to investigate whether psychotherapy trainees are more likely to accurately identify SPs if they were told about status differences beforehand compared to trainees who are ignorant of this fact.

### **Implications**

In conclusion, SPs can potentially profit greatly with regard to their authenticity through thorough training. Thus, we strongly recommend training SPs prior to implementation. We conclude that psychotherapy trainees (at an early stage of training) are not able to distinguish between SPs and real patients, indicating that high levels of authenticity can be reached. Considering the new psychotherapy law, it would be of particular interest to replicate the current results with psychology students.

### References

- Adamo, G. (2003). Simulated and standardized patients in OSCEs: Achievements and challenges 1992-2003. *Medical Teacher*, 25(3), 262–270. <https://doi.org/10.1080/0142159031000100300>
- Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, 21(1), 1–12. <http://dx.doi.org/10.1037/met0000051>
- Andrews, G., Henderson, S., & Hall, W. (2001). Prevalence, comorbidity, disability and service utilisation: Overview of the Australian National Mental Health Survey. *The British Journal of Psychiatry*, 178(2), 145–153.
- Angst, J., Gamma, A., Endrass, J., Goodwin, R., Ajdacic, V., Eich, D., & Rössler, W. (2004). Obsessive-compulsive severity spectrum in the community: Prevalence, comorbidity, and course. *European Archives of Psychiatry and Clinical Neuroscience*, 254(3), 156–164. <https://doi.org/10.1007/s00406-004-0459-4>
- Ariel, B., Sutherland, A., & Bland, M. (2019). The trick does not work if you have already seen the gorilla: How anticipatory effects contaminate pre-treatment measures in field experiments. *Journal of Experimental Criminology*, 55–66. <https://doi.org/10.1007/s11292-019-09399-6>
- Ay-Bryson, D. S., Weck, F., & Kühne, F. (2020). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*. <https://doi.org/10.1037/tep0000349>
- Ay-Bryson, D. S., Weck, F., & Kühne, F. (under review). Can students in simulation portray a psychotherapy patient authentically with a detailed role-script? Results of a randomized-controlled study. *Manuscript Submitted for Publication*.

- Barrows, H. S., & Abrahamson, S. (1964). The programmed patient: A technique for appraising student performance in clinical neurology. *Journal for Medical Education, 39*(8), 802–805.
- Brem, B., Christen, R., Bauer, D., Wüst, S., Schnabel, K., & Woermann, U. (2018). *Handbuch für SchauspielpatientInnen der Medizinischen Fakultät der Universität Bern*. Institut für Medizinische Lehre.
- Entwurf eines Gesetzes zur Reform der Psychotherapeutenausbildung, no. 19/9770 (2019).  
[https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3\\_Downloads/Gesetze\\_und\\_Verordnungen/GuV/P/Psychotherapeutenausbildung\\_Reform\\_Bundestag.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/P/Psychotherapeutenausbildung_Reform_Bundestag.pdf)
- Gräf, M., & Unkelbach, C. (2016). Halo effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy. *Personality and Social Psychology Bulletin, 42*(3), 290–310.  
<https://doi.org/10.1177/0146167215627137>
- Hautzinger, M. (2013). *Kognitive Verhaltenstherapie bei Depressionen: Mit Online-Materialien*. Beltz.
- Hodges, B. D., Hollenberg, E., McNaughton, N., Hanson, M. D., & Regehr, G. (2014). The psychiatry OSCE: A 20-year retrospective. *Academic Psychiatry, 38*(1), 26–34.
- Hoyt, W. T. (2002). Bias in participant ratings of psychotherapy process: An initial generalizability study. *Journal of Counseling Psychology, 49*(1), 35–46.  
<https://doi.org/10.1037/0022-0167.49.1.35>

- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.
- Krahn, L. E., Bostwick, J. M., Sutor, B., & Olsen, M. W. (2002). The challenge of empathy. *Academic Psychiatry*, *26*, 26–30.  
<https://doi.org/10.1176/appi.ap.26.1.26>
- Kühne, F., Maaß, U., & Weck, F. (2020). Einsatz standardisierter Patienten im Psychologiestudium: Von der Forschung in die Praxis. *Verhaltenstherapie*.  
<https://doi.org/10.1159/000509249>
- Kühne, F., Ay, D. S., Otterbeck, M., & Weck, F. (2018). Standardized patients in clinical psychology and psychotherapy: A scoping review on barriers and facilitators for implementation. *Academic Psychiatry*, *42*(6), 773–781.  
<https://doi.org/10.1007/s40596-018-0886-6>
- Kühne, F., Heinze, P. E., & Weck, F. (2020). Standardized patients in psychotherapy training and clinical supervision: Study protocol for a randomized controlled trial. *Trials*, *21*(276), 1–7. <https://doi.org/10.1186/s13063-020-4172-z>
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362.  
<https://doi.org/10.1177/1948550617697177>
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.
- Martell, R. F., & Evans, D. P. (2005). Source-monitoring training: Toward reducing rater expectancy effects in behavioral measurement. *Journal of Applied Psychology*, *90*(5), 956–963. <https://doi.org/10.1037/0021-9010.90.5.956>

- Melluish, S., Crossley, J., & Tweed, A. (2007). An evaluation of the use of simulated patient role-plays in the teaching and assessment of clinical consultation skills in clinical psychologists' training. *Psychology Learning & Teaching, 6*(2), 104–113.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9), S63-7. <https://doi.org/10.1097/00001888-199009000-00045>
- Mumma, G. H. (2002). Effects of three types of potentially biasing information on symptom severity judgments for major depressive episode. *Journal of Clinical Psychology, 58*(10), 1327–1345. <https://doi.org/10.1002/jclp.10046>
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review, 33*(3), 484–499. <https://doi.org/10.1016/j.cpr.2013.01.010>
- Partschefeld, E. (2013). *Evaluation des Einsatzes von Simulationspatienten in der psychotherapeutischen Ausbildung: Ein Beitrag zur empirischen Fundierung der Psychotherapieausbildung* [Doctoral dissertation].
- Partschefeld, E., Strauß, B., Geyer, M., & Philipp, S. (2013). Simulationspatienten in der Psychotherapieausbildung. *Psychotherapeut, 58*(5), 438–445. <https://doi.org/10.1007/s00278-013-1002-8>
- Peters, T., & Thrien, C. (2018). *Simulationspatienten: Handbuch für die Aus-und Weiterbildung in medizinischen-und Gesundheitsberufen*. Hogrefe. <https://doi.org/10.1024/85756-000>.
- Scherer, M., & Ehrhardt, M. (2017). *Handbuch für Simulationspatienten am Universitätsklinikum Hamburg- Eppendorf*. Institut für Allgemeinmedizin.

- Sharpless, B. A., & Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clinical Psychology Review, 29*(1), 47–56.
- Sheen, J., Sutherland-Smith, W., Thompson, E., Youssef, G. J., Dudley, A., King, R., Hall, K., Dowling, N., Gurtman, C., & McGillivray, J. A. (2020). Evaluating the impact of simulation-based education on clinical psychology students' confidence and clinical competence. *Clinical Psychologist*.
- Voderholzer, U. (Herausgeber). (2007). *Lehre im Fach Psychiatrie und Psychotherapie: Ein Handbuch*. Kohlhammer Verlag.
- Wirtz, M. A. (2017). Interrater Reliability. In *Encyclopedia of Personality and Individual Differences* (pp. 1–4). Springer International Publishing AG. 10.1007/978-3-319-28099-8\_1317-1
- Wünderlich, M. S., Nissen, C., Feige, B., Philipsen, A. S., & Voderholzer, U. (2012). Portrayal of psychiatric disorders: Are simulated patients authentic? *Academic Psychiatry, 36*(6), 501–502. <https://doi.org/10.1176/appi.ap.11090163>

**Table 1.** Descriptive results on APD and additional items' ratings.

Interviewed person number (status)	Authenticity	Real patient	Impairment
	<i>M (SD)</i>		
1 (SP)	1.96 (.43)	.36 (.22)	.40 (.21)
2 (real patient)	1.89 (.63)	.45 (.32)	.52 (.22)
3 (SP)	1.99 (.67)	.54 (.25)	.53 (.20)
4 (SP)	1.96 (.67)	.47 (.33)	.43 (.23)
5 (real patient)	2.13 (.59)	.65 (.28)	.40 (.18)
6 (SP)	2.48 (.45)	.69 (.24)	.62 (.21)
<b>Group</b>			
SPs	2.10 (.35)	.51 (.14)	.50 (.13)
Real patients	2.01 (.49)	.55 (.24)	.46 (.16)

*Note.* Authenticity was measured with the APD scale on a 4-point Likert scale: 0 = strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree; Items real patient and impairment were scored from 0 to 100%, and reported in the Table as normalized proportions; *M* = mean score; *SD* = standard deviation; SP = standardized patient.

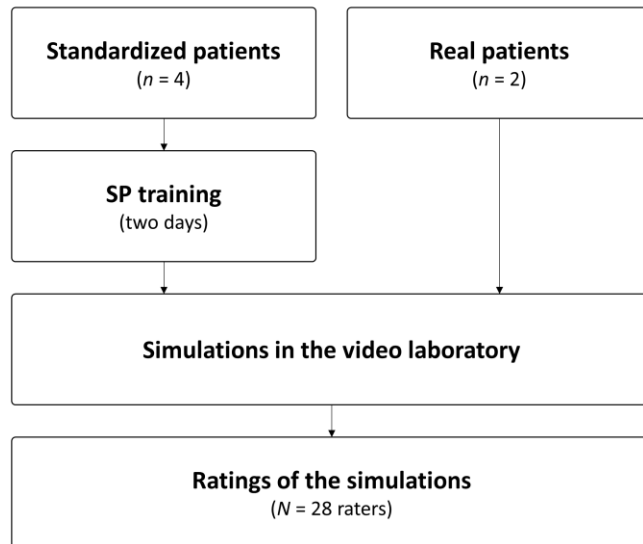
**Table 2.** Pearson's correlations based on ratings ( $N = 28$ ).

Measure	1	2	3	4	5	6
	(95% confidence interval)					
1 Authenticity RP	1					
2 Authenticity SP	.61** (.30, .80)	1				
3 Real patient RP	.76** (.53, .88)	.36 (-.02, .64)	1			
4 Real patient SP	.19 (-.19, .53)	.62** (.31, .80)	.27 (-.12, .58)	1		
5 Impairment RP	.13 (-.26, .48)	-.08 (-.44, .30)	.43 (.06, .69)	.10 (-.28, .46)	1	
6 Impairment SP	-.19 (-.53, .20)	.19 (-.20, .52)	-.19 (-.52, .20)	.52 (.18, .75)	.23 (-.16, .56)	1

*Note.* RP = real patient; SP = standardized patient.

*Bonferroni* corrected  $p$ -values: \*  $p < .003$ , \*\*  $p < .001$ .



**Figure 1***Study design*

## Supplement 1

### Rollenanleitung

**Titel:** Depressive Patientin

**Name, Geschlecht:** Frau Schwarz, weiblich

**Alter:** 19 Jahre

**Äußeres Erscheinungsbild:** Jungdliches Auftreten, wirkt verunsichert.

**Persönlichkeit:** Sie sind eher ruhig, aber haben gern Kontakt zu Menschen und sind gern in Gesellschaft. Sie sind spontan und planen nicht weit im Voraus. In letzter Zeit planen Sie eigentlich gar nichts mehr und verbringen den Großteil Ihrer Zeit zu Hause. Momentan haben Sie keinen Job und Ihre Aktivitäten haben in letzter Zeit stark nachgelassen. Sie sind unglücklich, weil Ihnen Ihre gewohnten Hobbies keine Freude mehr bereiten. Für Unternehmungen fehlen Ihnen Motivation und Antrieb. Sie sind unglaublich niedergeschlagen und grübeln nur noch vor sich hin.

**Aktuelle soziale und berufliche Situation:** Nach dem Abitur im letzten Jahr immer noch unentschlossen, wie es weitergeht; ledig; wohnt bei den Eltern; jobbt unregelmäßig.

**Ort des Gesprächs:** Therapiezimmer

**Dauer:** ca. 20 Minuten

**Situationsbeschreibung:** Es gab bereits Termine mit Ihrem Therapeuten, v.a. zur Diagnosestellung und zum Vereinbaren eines Behandlungsplans. Im aktuellen Gespräch geht es um Ihr starkes Grübeln, und darum, welche Gedanken Sie in konkreten, als negativ empfundenen Situationen belasten. Es geht außerdem darum, wie Sie mit solchen Gedanken konstruktiv umgehen können. Der Therapeut beginnt das Gespräch. Er kennt Ihren Namen und Ihr Alter.

**Aufgabe:** Der Therapeut soll mit Ihnen besprechen, welche belastenden Gedanken in konkreten Situationen auftreten, wie diese mit Ihrer Stimmung zusammenhängen und wie Sie sie besser erkennen können.

**Konkrete Situationen, in denen belastende Gedanken auftreten:** Wenn Sie darüber nachdenken, wie Ihre Zukunft aussehen könnte, denken Sie „Ich hab keine Ahnung, was ich machen soll.“; „Ich schaffe es sowieso nicht.“; „Ich bin zu nichts zu gebrauchen.“; „Ich weiß überhaupt nicht, was ich will.“; „Ich werde niemals etwas erreichen.“

Wenn Freunde Sie fragen, ob Sie etwas gemeinsam unternehmen wollen und Sie wiederholt absagen, denken Sie „Das bringt ja alles doch nichts.“; „Niemand interessiert sich wirklich für mich.“; „Das wird nie mehr besser.“; „Ich will niemandem zur Last fallen.“; „Ich möchte ihnen nicht ihre schöne Zeit versauen.“

Anfangs sind Sie von der unumstößlichen Echtheit dieser Gedanken überzeugt. Auf Bitten des Therapeuten, Belege für diese Gedanken anzuführen, tun Sie sich zunächst schwer. Auf Initiative des Therapeuten sind Sie offen für alternative Sichtweisen und das Überprüfen Ihrer Gedanken.

**Beispielsituation:** Während Sie mit Ihrem Hund draußen waren, trafen Sie zufällig einen Schulfreund. Er sprach Sie an und erzählte ungefragt ausgiebig von der Ausbildung, die er vor einigen Wochen begonnen hat. Sie fühlten sich augenblicklich beschämt und dachten, Sie werden nie so erfolgreich sein wie er. Als der Freund Sie nach Ihren aktuellen Lebensumständen fragte, wichen Sie aus Scham mit Ihrer Antwort aus. Aus Termingründen verabschiedete sich Ihr Freund, ohne eine weitere Nachfrage zu stellen. Sie dachten, dass sich niemand richtig für Sie interessiere. Während des Heimwegs grübelten Sie lange, dass Sie zu nichts zu gebrauchen seien und in der Zukunft sowieso nichts schaffen würden.

**Körpersprache:** Ihr Gesichtsausdruck ist niedergeschlagen und nachdenklich. Ihre Körperhaltung ist zusammengesunken.

### Krankengeschichte

**Hauptbeschwerde:** Nach dem Abitur im letzten Jahr haben Sie verschiedene Gelegenheitsjobs (bspw. in einer Bar, als Hausaufgabenbetreuung) angenommen, um Geld zu verdienen. Inzwischen haben sich die meisten Ihrer Freunde für eine Ausbildung oder ein Studium entschieden. Sie fühlen sich unter Druck gesetzt, sich ebenso für einen Karriereweg zu entscheiden. Jedoch fällt Ihnen diese Entscheidung sehr schwer. Sie schwanken zwischen einem Beruf mit Tieren (bspw. Tierpflegerin, Tiermedizin) oder einer Tätigkeit im sozialen Bereich (bspw. Lehramtsstudium). Sie fühlen sich orientierungslos und haben den Eindruck, zu nichts zu taugen. Sie sind niedergeschlagen und deprimiert. Sie machen sich große Sorgen, den Anschluss zu verlieren und keine Perspektive zu finden. Am meisten belastet Sie das viele Grübeln um Ihre Zukunft. Diesbezüglich sind Sie hoffnungslos, überzeugt davon nicht genug Willenskraft zu haben und nichts Gutes zu verdienen.

**Emotionale Situation:** Seit etwa sechs Monaten fühlen Sie sich deprimiert und hoffnungslos. In Ihrem Kopf kreisen die Gedanken ständig darum, dass Sie unfähig sind sich um Ihre Zukunft zu kümmern. Es fehlt Ihnen jegliche Motivation sich weiterhin nach Jobs umzusehen. Treffen mit Freunden aus der Schulzeit erleben Sie als demotivierend, da diese inzwischen studieren oder eine Ausbildung machen. Sie sind überzeugt, dass Ihre Freunde und Ihre Eltern Sie inzwischen als Versagerin sehen, die nichts auf die Reihe bekommt. Sie wünschten, Sie könnten etwas an der Situation ändern, doch Sie fühlen sich der Situation hilflos ausgeliefert.

Sie sind niedergeschlagen und antriebslos, es fällt Ihnen schwer, sich für alltägliche Aktivitäten wie Einkaufen oder Spaziergänge mit dem Hund zu motivieren.

**Soziale Situation:** Ihre Freunde, von denen einige noch in derselben Stadt wohnen, sehen Sie in den letzten Monaten immer seltener. Auch wenn Sie es ungern zugeben, haben Sie nur wenig Interesse an diesen Treffen. Sie haben den Eindruck, den Bezug zu Ihren Freunden zu verlieren.

Sie wohnen mit Ihrer jüngeren Schwester noch bei Ihren Eltern. Ihre Mutter kümmert sich um Haushalt, Wäsche und Essen. Sie glauben, Ihren Eltern zur Last zu fallen. Gesprächen rund um Ihre Zukunft gehen Sie aus dem Weg, Sie ziehen sich häufig in Ihr Zimmer zurück.

**Tagesablauf:** Da Sie aktuell nicht arbeiten, haben Sie keinen festen Tagesablauf. Sie gehen meist erst spät schlafen, verbringen den Tag mit Fernsehen und Computerspielen. Manchmal helfen Sie Ihrer jüngeren Schwester bei den Hausaufgaben oder gehen mit dem Hund spazieren. Abends essen Sie mit Ihrer Familie, ziehen sich aber schnell wieder zurück.

### Hintergrundinformationen

**Jetzige Beschwerden:** Sie leiden unter dem ständigen Grübeln über Ihre Zukunft. Diese Gedanken tragen zu Ihrer gedrückten Stimmung bei. Sie können sich zu nichts mehr motivieren.

**Sozialanamnese:** Sie haben eine Schwester, die 14 Jahre alt ist. Ihre Eltern sind beide Ende 50 und noch berufstätig. Das familiäre Verhältnis ist gut, insbesondere mit Ihrer Schwester verstehen Sie sich gut. Nach dem Abitur letztes Jahr fehlt es Ihnen an Orientierung. Sie haben den Eindruck, Ihre Familie erwarte von Ihnen eine Entscheidung, wie Ihre Karriere verlaufen soll.

**Körperliche Anamnese:** Bisher keine ernsthaften körperlichen Erkrankungen. Keine Medikamente, kein Substanzgebrauch.

**Vorgeschichte:** Sie waren in Ihrem Leben bisher noch nie beim Psychotherapeuten oder Psychiater. Sie hatten noch nie eine psychische Erkrankung (keine Depression, keine Manien etc.). Sie leiden nicht unter Suizidalität.

## 7.4 Eidesstattliche Erklärung

Hiermit erkläre ich, Destina Sevde Ay-Bryson (geb. Ay, 20.03.1992 in Berlin), dass ich die vorliegende Dissertation "The authenticity of simulated patients in psychotherapy training and research" selbstständig und ohne Hilfe Dritter verfasst habe und bei der Abfassung alle Regelungen guter wissenschaftlicher Standards eingehalten wurden. Ich erkläre, dass die Dissertation in der gegenwärtigen Fassung keiner anderen Hochschule zur Begutachtung vorgelegen hat oder vorliegt. Ich erkläre, an keiner anderen Hochschule ein Promotionsverfahren eröffnet zu haben.

Potsdam, 17. Mai 2021

.....  
Destina S. Ay-Bryson