INTRODUCTION

In recent years, the number of patients in need of more demanding nursing care in acute care hospitals has increased (Dharmarajan et al., 2016; Smith et al., 2018). Additionally, a decrease in length of stay has been observed in many countries, leading to more complex nursing care situations on the hospital unit level (Bartholomeyczik, 2011; Borghans, Heijink, Kool, Lagoe, & Westert, 2008; Hamada, Sekimoto, & Imanaka, 2012). This may jeopardize patient safety without suitable staffing (Aiken et al., 2017; Driscoll et al., 2018). For example, a recent study showed higher probability of death and length of stay and lower probability of being discharged home in internal medical patients screened as experiencing highly complex care episodes during their hospital stay (Bandini et al., 2018). Moreover, Registered Nurses (RN) at acute care hospitals felt overwhelmed in highly complex nursing situations if contextual and individual resources were...
not supportive (Kentischer, Kleinknecht-Dolf, Spirig, Frei, & Huber, 2018). This shows the need for an instrument for monitoring complexity of nursing care on the patient level in acute care hospitals to allocate adequate nursing staff resources according to grades of patient complexity, thus safeguarding high patient safety levels and keeping RNs motivated.

1.1 | Background

Complexity has been defined as different dynamically interacting components causing uncertainty, unpredictability and the need to find unique solutions (Kannampallil, Schauer, Cohen, & Patel, 2011; Lindberg & Lindberg, 2008). RNs require professional knowledge, skills, experience, attention, clinical reasoning abilities and supportive contextual conditions to be able to handle complex nursing care situations successfully (Hoeve, Brouwer, Roodbol, & Kunnen, 2018; Kentischer et al., 2018; Kuiper, O'Donnell, Pesut, & Turrisi, 2017).

When we started developing our instrument in 2009, various attempts at defining and measuring complexity of nursing care had been made. Definitions found in the literature covered nurses’ knowledge and skill levels about procedures and nursing care processes (Prescott & Phillips, 1988), patients’ needs about education, decision-making and procedures (Fulton & Wilden, 1998), patients’ nursing and medical conditions and the characteristics of nurses and care environments (Petryshen, Pallas, & Shamian, 1995) and nurses’ education levels to generate changes and set priorities in patients’ care (Blastorah et al., 2001).

Of existing instruments, some measured complexity using instruments that also addressed other concepts (Berry, 1977; Fulton & Wilden, 1998; Prescott & Phillips, 1988). One instrument focussed on patients’ needs rather than on demands on nursing care (Blastorah et al., 2001). Another followed an interprofessional approach (de Jonge et al., 2001).

Various nursing researchers developed their instruments relying on the “framework for the comparative analysis of the complexity of organizations” described by Perrow (1967). They assessed complexity for ambulatory care (Verran & Reid, 1987), nursing home care (Velasquez, 2007) or on the hospital unit level rather than on the patient level (Alexander & Kroposki, 2001; Leatt & Schneck, 1981; Overton, Schneck, & Hazlett, 1977). According to this framework, organizations differ about work complexity by the stability and understanding of the raw materials used and by the predictabilities or uncertainties concerning the technology workers apply to change the raw materials according to the organization’s aims. In nursing, raw materials refer to patients and the stability of their health situations. Technology means the interventions RNs practice relating to the nursing care process. Additionally, work complexity diverges about variability in raw materials and technology, that is according to variability among patients and their relatives, health problems and nursing interventions. This framework defines complexity as a relational concept incorporating a patient’s situation and the resultant demands on nursing care (Alexander & Kroposki, 2001; Kleinknecht-Dolf, Grand, et al., 2015; Overton et al., 1977; Perrow, 1967).

However, these instruments were not congruent with the purpose of our monitoring. Accordingly, we developed a new German instrument following a framework for instrument development with the intersecting elements of an instrument’s function, its content, its structure and its measurement model (Brühl, 2012). About its function, our instrument “Complexity of Nursing Care” is to be used for monitoring purposes and to differentiate between patients, wards and medical disciplines by means of quantifying complexity in one reliable value (Kleinknecht-Dolf, Grand, et al., 2015).

About the content, we relied on Perrow’s (1967) framework, given its previous usefulness in nursing and composed this working definition: “Complexity of nursing care describes the demands on professional nursing in the acute care hospital setting. The extent of the complexity is determined by the degree of instability, variability and uncertainty that exists. These factors are influenced in part by the impact of the disease and its associated therapies on the patient and on the patient’s everyday life. However, they are also influenced by decisions about care and interventions taken to deliver the best outcome for the patient” (Kleinknecht-Dolf, Grand, et al., 2015, pp. 592–593).

An instrument’s structural model describes the relationships between latent variables, hence between constructs and sub-constructs that are not directly measurable such as complexity and its dimensions. The measurement model clarifies how the latent variable(s) are operationalized. Relationships are called reflective when measurement variables are effects of a latent variable and formative when they are their causes (Brühl, 2012; Hair, Hult, Ringle, & Sarstedt, 2017). The structural and measurement model of the first version of our instrument included 15 formative items (measurement variables) on three subscales (latent variables).

Subscale One explored the complexity of the patient situation, Subscale Two the demands on nursing care and Subscale Three the required competencies of RNs. Clinical nurse specialists (CNS) discussed the instrument for face validity and congruence with Perrow’s (1967) framework before it was investigated in a pilot study (Kleinknecht-Dolf, Grand, et al., 2015), followed by a sequential explanatory mixed-methods study between 2010–2014 (Kleinknecht-Dolf, Spichiger, et al., 2015). Cluster and regression analyses and focus group interviews favoured an instrument with seven items, collapsing some of the items, deleting items not statistically relevant and deleting Subscale Three (Guggenbühl, Hausmann, & Müller, 2012).

This advanced instrument is reevaluated with this study. Our aims were to psychometrically test the instrument “Complexity of Nursing Care”, to represent complexity in one score and to broaden the understanding of the instrument’s psychometrics and applicability.

2 | THE STUDY

2.1 | Design

We applied an embedded mixed-methods design with a leading quantitative study section (Creswell, 2014). The research protocol for this study was published elsewhere (Huber, Kleinknecht-Dolf, Müller, Kugler, & Spirig, 2016).
2.2 | Methods

2.2.1 | Instrument

Based on Perrow’s (1967) framework and the preceding results, we hypothesized a second-order measurement and structural model with the latent variables “patient situation” and “demands on nursing care” as first level and “complexity of nursing care” as second level and formative relationships between measurement and latent variables as shown in Figure 1 (Hair et al., 2017).

“Patient situation” was operationalized with: (a) clinical signs and symptoms; (b) risks; and (c) conditions of psychosocial burden. “Demands on nursing care” was measured with: (d) self-care abilities; (e) decision-making; (f) adaptation of nursing care; and (g) predictability of the impact of nursing interventions. The items were specified with a title, a description and examples of its meaning and contain a five-point Likert scale with an individual verbal description for every measurement point (Appendix S1). Furthermore, we provided examples for every measurement point in an additional form in four versions: for adult patients, paediatric patients, maternity wards and neonatal wards. Additionally, we added one global item for rating overall demands of the patient’s care on RNs on a 10-point rating scale (1 = extremely low, 10 = extremely high).

2.2.2 | Quantitative data collection

The quantitative study section involved three Swiss university and two Swiss tertiary hospitals. In the prior pilot study, we tested the instrument “Complexity of Nursing Care” on six wards of two hospitals between April–May 2015. All new stationary patients were assessed daily by the RN responsible for the day shift for up to 5 days over a 2-week period. Some assessments were repeated by the head nurse, the CNS, or another RN on the ward for the estimation of the inter-rater reliability.

In the ensuing main cross-sectional study, all stationary patients were assessed over 4 weeks by the RN responsible for the day shift in all hospitals for up to 5 days starting from admission in November–December 2015 with the instrument “Complexity of Nursing Care,” added to the electronic documentation system in almost all wards.

2.2.3 | Qualitative data collection

The qualitative study section aimed to explore how complexity crystallized in various nursing care situations and how the instrument was applied in these situations. We performed 12 collective case studies between August–November 2015 (Creswell & Poth, 2018; Yin, 2014). Four nursing care situations were chosen in each of the following departments: (a) internal medicine; (b) cardiology, angiology and pneumatology; and (c) gynaecology. Pre-defined criteria per department included a patient with low complexity, one with high complexity, one with a hospital stay of 1–3 days and one with a hospital stay of at least 1 week.

For every situation, a researcher analysed the patient's record about information pertaining to our working definition on one specific day. On the same day, the RN primarily responsible for the patient and the ward’s CNS assessed the situation independently with the instrument “Complexity of Nursing Care”. Afterwards, they were interviewed in focussed single interviews (Flick, 2017). The interviews covered experiences while applying the instrument, considerations on how the situation was rated and perceptions of the patient’s demands on nursing care. RNs employed for less than three months on the ward were not included.

2.2.4 | Syntheses

For the syntheses, the samples from the quantitative the qualitative study section were used.

2.3 | Data analysis

2.3.1 | Quantitative analyses

Inter-rater reliability

Double assessments from the pilot study were analysed on the item level using Spearman’s rank correlation.
Model refinement

The preliminary hypothesized structural and measurement model was reviewed by calculating Gamma coefficients ($\gamma$) between the measurement variables and network analysis based on partial correlations to obtain information of the relationships among the measurement variables and thereby identify the actual model of the instrument to be evaluated (Costantini et al., 2015).

Evaluation of the structural and measurement model

We used partial least square structural equation modelling (PLS-SEM) for the evaluation of the refined structural and measurement model and the estimation of the instrument’s psychometrics. PLS-SEM is an explorative modelling approach that aims to predict a target construct incorporating reflective and formative measures and thus to confirm construct validity of an instrument.

PLS-SEM is a non-parametric scaling method that does not require normal distribution of the data. Reflective models are evaluated by calculating indicator reliability and internal consistency of the measurement variables, convergent validity in terms of shared variance between the measurement variables of a reflective latent variable and discriminant validity defined as sufficient distinction between reflective latent variables. In formative models, collinearity and path coefficients are calculated to investigate whether the formative indicators do not measure similarly and that each indicator contributes sufficiently to the latent variable (Hair et al., 2017). The global item was added to estimate convergent validity of the formative model.

Statistical analyses were performed with IBM SPSS Statistics 23 (SPSS Inc, Released 2015) and the statistical computing language R 3.1.3 (R Foundation for Statistical Computing, 2016).

2.3.2 | Qualitative analyses

Within-case analysis

We performed within-case analysis inductively using summarizing and explicating content analysis before constructing a case narrative for each case (Mayring, 2015; Patton, 2015; Yin, 2014). MAXQDA12 was used in this step (MAXQDA, 1989–2016).

Cross-case analysis

Cross-case analysis was completed using a table applying structuring content analysis to explore similarities and differences among the cases’ contents representing the dimensions instability, uncertainty and variability (Creswell & Poth, 2018; Mayring, 2015; Yin, 2014). Additionally, we compared the individual interviews about characteristics of different complexity levels. Statements concerning the instrument’s applicability were analysed with summarizing content analysis (Mayring, 2015).

2.3.3 | Syntheses

We synthesized quantitative and qualitative results in three steps (Creswell, 2015; Fetters, Curry, & Creswell, 2013). First, inter-rater reliability scores of the pilot study were contrasted in a table to those of the case studies and to narrative agreements and disagreements of the two interview participants of the same case. Second, in a discussion, the instrument’s psychometrics were supplemented with the interview participants’ experiences while applying the instrument. Third, the meaning of the complexity score in practice was explored by supplementing it with the cross-case results in a table.

2.4 | Ethics

A cantonal ethics committee assessed this research programme, which did not fall under Swiss human research legislation, as ethically unproblematic (Waiver No. 82/14 from 5 December 2014 and Waiver No. 49 from 22 May 2015). We followed Swiss legislation on data protection and guidelines for good clinical practice. All patient identification numbers and names of interview participants were replaced with codes. Participation in the interviews was voluntary. Interview participants signed the informed consent form after being given written and verbal study information by one of the researchers. All RNs involved in the quantitative data collection were given verbal and written study information and on-going support on the wards from a member of the research team. According to Swiss law, there was no requirement for asking consent from the patients or from RNs participating in the quantitative data collection. Data were stored safely.

2.5 | Rigour

We used the same theoretical understanding in every analytical step (Creswell, 2014). An expert statistician processed the statistical analyses. The first author reflected on analytical steps and results regularly with research team members and in peer groups (Rettke, Pretto, Spichiger, Frei, & Spirig, 2018). Further, we sent the case study results to the interview participants and three research team members for validation in Spring 2017 leading to adjustment of the results (Mayring, 2015).

3 | RESULTS

3.1 | Quantitative study section

In the pilot study, 748 assessments were performed on 246 patients. From these, 60 assessments of 47 patients were repeated by a second-rater. In the main study, 10,892 patients were assessed with 32,610 assessments on 189 wards to investigate the instrument in a broad variety of medical disciplines.

3.1.1 | Inter-rater reliability

The ratings ($N = 60$) of two pairs of raters showed moderate to strong positive correlations for all of the items with Spearman’s Rho ($r_s$) between .703–.855 (Table 3).
3.1.2 | Model refinement

We included every patient’s first assessment in the pilot study \((N = 246)\) to refine the structural and measurement model. The first assessments of the patients of one university hospital were selected in the main study \((N = 2,412)\) to review this refined model.

Items 1 and 2, items 4 and 5 and items 6 and 7 strongly correlated among each other in the pilot and the main study. Other correlations were slightly weaker (Table 1). Thus, these groups were assumed to build three formative indicators of the latent construct “complexity of nursing care,” while item three proved to be an indicator of its own. Network analyses confirmed this model in the pilot and the main study. Accordingly, a reflective-formative second-order model was developed in the pilot study and was confirmed in the main study (Figure 2).

3.1.3 | Evaluation of the structural and measurement model

In the main study \((N = 2,412)\), the reflective measurement model was supported through satisfactory estimates of:

**TABLE 1** Inter-item correlations based on gamma-coefficients, main study \((N = 2,412)\)

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
<th>Item 6</th>
<th>Item 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>1.00</td>
<td>.76</td>
<td>.56</td>
<td>.68</td>
<td>.58</td>
<td>.69</td>
<td>.69</td>
</tr>
<tr>
<td>Item 2</td>
<td>.76</td>
<td>1.00</td>
<td>.64</td>
<td>.69</td>
<td>.64</td>
<td>.72</td>
<td>.76</td>
</tr>
<tr>
<td>Item 3</td>
<td>.56</td>
<td>.64</td>
<td>1.00</td>
<td>.53</td>
<td>.58</td>
<td>.59</td>
<td>.62</td>
</tr>
<tr>
<td>Item 4</td>
<td>.68</td>
<td>.69</td>
<td>.53</td>
<td>1.00</td>
<td>.79</td>
<td>.74</td>
<td>.67</td>
</tr>
<tr>
<td>Item 5</td>
<td>.58</td>
<td>.64</td>
<td>.58</td>
<td>.79</td>
<td>1.00</td>
<td>.71</td>
<td>.68</td>
</tr>
<tr>
<td>Item 6</td>
<td>.69</td>
<td>.72</td>
<td>.59</td>
<td>.74</td>
<td>.71</td>
<td>1.00</td>
<td>.78</td>
</tr>
<tr>
<td>Item 7</td>
<td>.69</td>
<td>.76</td>
<td>.62</td>
<td>.67</td>
<td>.68</td>
<td>.78</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**FIGURE 2** Reflective-formative second-order model
† Loadings of the reflective measurement model of the main study
‡ Path coefficients of the formative structural model of the main study

**TABLE 2** Discriminant validity, main study \((N = 2,412)\)

<table>
<thead>
<tr>
<th>Item</th>
<th>First-order latent variable</th>
<th>Patient status</th>
<th>Psycho-social burden</th>
<th>Patient’s abilities</th>
<th>Nursing care process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Signs and symptoms</td>
<td>Patient status</td>
<td>0.91</td>
<td>0.52</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td>2 Risks</td>
<td>Patient status</td>
<td>0.92</td>
<td>0.58</td>
<td>0.67</td>
<td>0.70</td>
</tr>
<tr>
<td>3 Psychosocial burden</td>
<td>Psychosocial burden</td>
<td>0.60</td>
<td>1.00</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>4 Self-care abilities</td>
<td>Patient’s abilities</td>
<td>0.71</td>
<td>0.52</td>
<td>0.94</td>
<td>0.69</td>
</tr>
<tr>
<td>5 Decision-making</td>
<td>Patient’s abilities</td>
<td>0.61</td>
<td>0.51</td>
<td>0.93</td>
<td>0.67</td>
</tr>
<tr>
<td>6 Adaptation nursing care</td>
<td>Nursing care process</td>
<td>0.68</td>
<td>0.54</td>
<td>0.70</td>
<td>0.92</td>
</tr>
<tr>
<td>7 Predictability of impact</td>
<td>Nursing care process</td>
<td>0.66</td>
<td>0.53</td>
<td>0.64</td>
<td>0.91</td>
</tr>
</tbody>
</table>
### TABLE 3 Side-by-side matrix inter-rater reliability

<table>
<thead>
<tr>
<th>Item</th>
<th>( r_s ) pilot study</th>
<th>( r_s ) case studies</th>
<th>Summary of the ratings of the case studies</th>
<th>Examples of narrative arguments with numeric disagreements</th>
<th>Examples of narrative arguments with numeric agreements</th>
<th>Interpretation</th>
</tr>
</thead>
</table>
| Clinical signs and symptoms         | .713                   | .856                    | 7 cases with identical ratings, 5 cases with differences between 0.5-1 point | **Consistent argumentation:** Case 3, RN (4) and CNS (5): multiple progressive signs and symptoms  
**Non-consistent argumentation:** Case 2, RN (3): multiple, severe signs and symptoms, to some degree controllable, but always a topic of discussion. CNS (4): stable vital signs, severe, insufficiently controlled pain | **Consistent argumentation:** Case 6, RN (2) and CNS (2): multiple, decreasing signs and symptoms.  
**Non-consistent argumentation:** Case 9, RN (2): no discomfort at rest in spite of instable vital signs. CNS (2): typical clinical signs, familiar knowledge needed on basic education level | Item 1 shows good inter-rater reliability despite certain subjectivity in the perception of the situation and some disagreement of how to rate the item |
| Risks                               | .758                   | .555                    | 4 cases with identical ratings, 7 cases with a 1-point difference and 1 case with a 2-point difference | **Consistent argumentation:** Case 5, RN (3) and CNS (4): multiple moderate to high risks  
**Non-consistent argumentation:** Case 7, RN (2): one risk to be controlled with an intervention being started. Possibility of hidden risks. CNS (4): multiple risks to monitor | **Consistent argumentation:** Case 3, RN (5) and CNS (5): multiple, extremely severe risks with a risk of lethal deterioration.  
**Non-consistent argumentation:** Case 1, RN (2): risks because of overweight, CNS (2): multiple risks because of actual primary disease and psychiatric co-morbidities | Item 2 shows good inter-rater reliability in the pilot study. In the case studies, some cases indicate no consistent understanding of the item or of the situation and some disagreement of how to rate the item |
| Conditions of psychosocial burden   | .772                   | .789                    | 4 cases with identical ratings, 7 cases with differences between 0.5-1 point, 1 case with one missing | **Consistent argumentation:** Case 12, RN (4) and CNS (5): extremely burdened family with two family members on intensive care units  
**Non-consistent argumentation:** Case 7, RN (2): patient shows no signs of insecurity. CNS (3): patient with personal resources and some signs of insecurity | **Consistent argumentation:** Case 4, RN (2) and CNS (2): patient initially exhausted.  
Case 6, RN (2) and CNS (2): patient with personal and familial resources | Item 3 shows good inter-rater reliability despite certain subjectivity in the perception of the situation and some disagreement of how to rate the item |
| Self-care abilities                 | .854                   | .781                    | 6 cases with identical ratings, 4 cases with a 1-point difference, 1 case with a 2-point difference and 1 person who was undecided | **Consistent argumentation:** Case 2, RN (4) and CNS (5): extensive instructions and extensive support of ambulation needed, patient able to eat independently  
**Non-consistent argumentation:** Case 9, RN (4): patient needs support and education. CNS (2): independent patient with unclear familiar support about his chronic conditions | **Consistent argumentation:** Case 10, RN (5) and CNS (5): very passive patient.  
**Non-consistent argumentation:** Case 1, RN (3): patient independent in personal hygiene, dependent in signing papers. CNS (3): patient partly dependent in personal hygiene with urinary incontinence and ambulation | Item 4 shows good inter-rater reliability despite certain subjectivity in the perception of the situation and the argumentation and some disagreement on how to rate the item |
<table>
<thead>
<tr>
<th>Item</th>
<th>( r_s ) pilot study</th>
<th>( r_s ) case studies</th>
<th>Summary of the ratings of the case studies</th>
<th>Examples of narrative argumentations with numeric disagreements</th>
<th>Examples of narrative argumentations with numeric agreements</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision-making</td>
<td>.758</td>
<td>.866</td>
<td>6 cases with identical ratings, 5 cases with a 1-point difference, 1 person did not rate this item</td>
<td>Non-consistent argumentation: Case 11, RN (3): patient not capable of making decisions, relative needs support. CNS (2) patient not capable of making decisions</td>
<td>Consistent argumentation: Case 10, RN (5) and CNS (5): patient not capable of making decisions, relative decides against professional recommendations. Non-consistent argumentation: Case 5, RN (3): patient knows her needs and concerns well, needs information for decisions. CNS (3): patient does not speak up for herself</td>
<td>Item 5 shows good inter-rater reliability despite certain subjectivity in the perception of the situation and the argumentation and some uncertainty of how to rate the item</td>
</tr>
<tr>
<td>Adaptation of nursing care</td>
<td>.690</td>
<td>.683</td>
<td>6 cases with identical ratings, 4 cases with a 1-point difference and 2 cases with a 2 point difference</td>
<td>Consistent argumentation: Case 10, RN (2): situation more or less stable. CNS (3): situation assessable and controllable, no sudden deterioration expected. Non-consistent argumentation: Case 8, RN (1): the actual day only interventions realized that had been planned the day before. CNS (3): constant re-evaluations needed because of the symptoms with some routinized adaptations</td>
<td>Consistent argumentation: Case 4, RN (1) and CNS (1): only routine interventions and monitoring. Case 12, RN (5) and CNS (5): constant adaptation because of high instability. Non-consistent argumentation: Case 11, RN (3): interventions have to be adapted to the patient's condition. CNS (3): risks have to be evaluated daily</td>
<td>Item 6 shows acceptable inter-rater reliability, however, there is no consistent understanding of this item in some cases and some disagreement on how to rate the item. One person did not understand this item at all</td>
</tr>
<tr>
<td>Predictability of the impact of nursing interventions</td>
<td>.713</td>
<td>.585</td>
<td>7 cases with identical ratings, 4 cases with a 1-point difference and 1 case with a 3-point difference</td>
<td>Non-consistent argumentation: Case 7, RN (2): stabilized situation because of interventions taken. CNS (3): we know the direction, how the situation will develop, which side effects are to be expected, but we do not know exactly what will happen</td>
<td>Non-consistent argumentation: Case 5, RN (3): this morning's interventions were effective. CNS (3): situation generally predictable, but still not too predictable because of the patient's multi-morbidity. Case 8, RN (2): interventions about own practice are predictable; interventions outside own practice are not predictable. CNS (2): impacts of medications and conversations are mostly predictable</td>
<td>Item 7 shows rather a good inter-rater reliability in the pilot study. In the case studies, some cases indicate no consistent understanding of this item and some disagreement on how to rate the item</td>
</tr>
</tbody>
</table>

Abbreviation: \( r_s \), Spearman’s rho.

The number indicates the rating on the scale between 1–5 points, where every measurement point of the scale has an individual verbal description.
• indicator reliability: all first-order latent variables explained sufficient variance of the items (loadings reported in Figure 2). To represent the measurement model for complexity, the items were assigned to complexity in the form of the repeated indicators approach (loadings: item 1 = 0.79; item 2 = 0.84; item 3 = 0.73; item 4 = 0.85; item 5 = 0.81; item 6 = 0.84; item 7 = 0.81);
• internal consistency: Cronbach’s alpha of items 1 and 2 about “patient status” was .80, of items 4 and 5 for “patient’s abilities” .86 and of items 6 and 7 for “nursing care process” .81; Cronbach’s alpha of all the seven items according to “complexity of nursing care” was .91;
• convergent validity: average variance extracted (AVE) confirmed convergence of the items about the suitable latent first-order variables with AVE-values of 0.83 for “patient status,” 0.88 for “patient’s abilities” and 0.84 for “nursing care process,”
• discriminant validity: four confined factors were approved (values reported in Table 2).

The latent variable “Conditions of psychosocial burden” contains only one item; thus, Cronbach’s alpha and AVE equaled 1.00. Evaluations of the formative structural model confirmed the proposed model with estimates of:

• variance inflation factors (VIF) supporting the first-order latent variables to be independent formative indicators according to the second-order latent variable “complexity of nursing care” (VIF-value for “patient status”: 2.64, for “conditions of psychosocial burden”: 1.71, for “participation”: 2.50 and for “nursing care process”: 2.75);
• statistically significant path coefficients for all four first-order latent variables (weights reported in Figure 2) and for the global item (path coefficient: .7791).

Based on these results, a complexity score (the linear combination of the values of the seven items with multiplying factors derived from the measurement and structural model) can be calculated. This is in conjunction with a scaling factor to get complexity scores between 1.0–5.0, where 1.0 equals lowest complexity and 5.0 corresponds to highest complexity. Descriptive analyses, including all five hospitals, showed a right-skewed distribution of the complexity scores (first hospitalization day (1st hd), N = 10,363: mean (m) = 2.118, standard deviation (SD) = 0.814; 2nd hd, N = 8,460: m = 2.145, SD = 0.807; 3rd hd, N = 6,379: m = 2.166, SD = 0.812; 4th hd, N = 4,435: m = 2.231, SD = 0.806; 5th hd, N = 2,545: m = 2.271, SD = 0.810).

3.2 | Qualitative study section

The participants were between 24–51 years old (mean: 37 years) and had numerous years of professional experience (6 participants >20 years, 4 participants 11–15 years, 8 participants 6–10 years and 3 participants 1–5 years). The average interview length was 49 min (range: 26–72 min). Three CNSs participated in two case studies.

3.2.1 | Applicability of the instrument

Most interviewees had to become familiarized with the instrument before being able to apply it: “I had to read the questionnaire twice. I had to read it carefully. And while completing it, I still had some difficulties (...) even though there are very good examples, it is still somehow subjective,” C10RN. Only one CNS judged the instrument as “easily understandable.” Multiple patient problems with different severity levels, unstable fluctuating situations, psychosocial issues, patients who were unable to communicate, or autonomous patients in need of information caused uncertainty. Items 6 and 7 caused the most problems, either because participants did not understand their meaning or because of the need for constant adaptations in highly unstable situations. Overall, the interview participants were able to explain their ratings.

3.2.2 | Within- and cross-case results

Two examples of central characteristics of case narratives are shown in Table 4. Further case narratives have been published previously (Gurtner, Spirig, Staudacher, & Huber, 2018).

In cross-case analysis, patient-related complexity of nursing care in acute care hospitals was confirmed as dynamic phenomenon between low and high complexity in the dimension’s instability, variability and uncertainty: “Slightly complex is close to complex. It can rapidly topple to the other side if the patient develops some complications,” C9CNS. The extent of complexity appeared:

• in the type, number, assessability, controllability, progress and threat of nursing-relevant illness-related, therapy-related, psychosocial and ethical problems of patients and their relatives as well as the interactions between them: “the assessment is very high because of the sum of his risks,” C12RN
• in the degree of concentrated attention, knowledge, experience and caring required from RNs to achieve an effective nursing care process for strengthening abilities and preventing deterioration, persistent harm and suffering of patients and for strengthening abilities and relieving distress of patients’ relatives and thus delivering the best outcomes for patients and their relatives: “I sat next to him and observed him – I even almost studied him,” C11RN.

Personal, cognitive, physical, functional, communicative and social resources of patients and their relatives were constitutive for the extent of complexity: “He has a long list of problems. But we have the impression that he is not suffering (...). He has a positive attitude; he is very thankful and patient,” C6CNS.

The comparison of the interviews in ascending order concerning the theoretical dimensions instability, uncertainty and variability showed that low complexity was characterized as “nursing care situations requiring attention in routine nursing care interventions for patients with high resources in controlled, stable conditions.” In contrast, high complexity was described as “nursing care situations requiring constant attention on severe, poorly assessable
illness-related and psychosocial problems of patients in uncontrolled unstable conditions, with uncertain outcomes and limited or in some cases missing resources" (Table 4).

### 3.3 Syntheses

#### 3.3.1 Inter-rater reliability and considerations of two raters

Items 1, 3, 4 and 5 showed comparable values of inter-rater reliabilities in the pilot study and the case studies ($r$: .718–.866). The disagreements between two interview participants were minor in these items.

The ratings of items 2 and 7 diverged in the pilot study ($r$: .757 respective .728) and the case studies ($r$: .555 respective .585). In item 6, the values of the pilot study and the case studies were rather low ($r$: .703 respective .683). The narrative arguments revealed divergent understandings of how these items should be rated (i.e. case 10, item 7, RN: “today, nothing significant will happen,” rated as “impacts and consequences of nursing interventions are predictable,” CNS: “we do not know where this whole situation is supposed to end,” rated as “short-term impacts and consequences are partly predictable, medium-term are mostly not predictable.”

In the case studies, some pairs of raters reasoned the same way but chose another rating in most of the items. Some raters reasoned differently about the same item but still arrived at an identical numerical rating. Finally, there was also consensus according to numerical rating and narrative explanations (Table 3).

### 3.3.2 Construct validity and applicability

The psychometric evaluations of the instrument supported its construct validity and reliability with all items yielding comparably good values. However, the interviews revealed that items 6 and 7 in particular seemed to be more difficult to rate than the others. Furthermore, the application of this instrument needed reflection. It can be hypothesized that these uncertainties diminish if raters become more familiar with the instrument, or that the amount of quantitative data equalizes some uncertainties.

### 3.3.3 Distribution and description of the complexity score

PLS-SEM led to a formula for the calculation of a continuous complexity score. Merging the complexity scores with cross-case results indicated that only a lower ratio of nursing care situations demanded high levels of attention, knowledge, skills and caring expertise to handle conditions of patients with limited abilities and multiple, instable, poorly controllable health, psychosocial and ethical problems (Table 4).

### 4 DISCUSSION

With this study, we developed a new instrument with satisfactory psychometric properties. It was evaluated using an adequate number of assessments in a broad variety of acute care hospital disciplines. This instrument makes it possible to capture the complexity...
of a nursing care situation in one reliable value. With only seven items, not much effort is needed to use this instrument in practice or research.

In the leading quantitative study section, we elaborated and evaluated a reflective-formative second-order model. All indices supported data-model fit and thus reliability and construct validity of the instrument sufficiently (Hair et al., 2017). Considering the first version of this instrument contained only formative indicators, this is an important advancement. With the former version, it was not possible to calculate a reliable composite score and to use the data for analyses other than descriptive evaluations based on ordinal increments. The advanced instrument allows us to work with the data by means of a continuous composite score and thus combine them with other monitoring data for more complex evaluations.

Interestingly, our preliminary operationalization of the sub-construct “patient situation” with the items “clinical signs and symptoms,” “risks” and “conditions of psychosocial burden” was differentiated in such a way that “clinical signs and symptoms” and “risks” refer to the present clinical state of a patient and “conditions of psychosocial burden” point to a more holistic knowledge of the patient and his or her family. Also, the operationalization of the “demands on nursing care” appeared to be composed of two sub-constructs. “Patients abilities” describes demands on nursing care if patients’ abilities are limited, while “nursing care process” represents the technical aspect of nursing care. This advanced our conceptual understanding based on Perrow’s (1967) framework.

Additionally, the qualitative study section added valuable information to the instrument and the theoretical framework. In particular, the sub-construct “patient’s abilities” illustrates the relational understanding of the complexity of nursing care. “Self-care abilities” and “decision-making” of patients or patients’ relatives relieve nurses about “doing for” and “enabling” patients with limited abilities, two characteristics of the concept of caring (Swanson & Wojnar, 2004). Abilities of self-management or decision-making are examples of internal resources, which help to effectively adopt external resources like healthcare services or social support (Steverink, Lindenberg, & Slaets, 2005). Interview participants incorporated additional resources of patients and their relatives. For instance, familial social support reduced or, if lacking, increased the participant’s ratings of psychosocial burden. Similarly, other authors described patients’ internal and external resources as defining elements of patient complexity (Schank et al., 2012; Shippee, Shah, May, Mair, & Montori, 2012).

Synthesis of the quantitative and qualitative study strands supports overvalue of an embedded mixed-methods design in instrument development. In particular, the conjunction of the quantitative measurements of inter-rater reliability with an analysis of considerations of two different raters yielded helpful explanations for marginally acceptable statistical values. This led us to rephrase the wording of some of the descriptions and examples in the instrument. Furthermore, the instrument’s robustness has to be questioned in items where pairs of raters used different ratings while indicating the same arguments. This initiated revisions of the verbal descriptions of the measurement points of some of the items. Also, the interview participants’ answers concerning the instrument’s applicability impelled improvements in the wording. With these advancements, we hope to further increase the instrument’s inter-rater reliability and help nurses to apply our new instrument with greater security.

Furthermore, our aim was representing and interpreting the complexity of nursing care with one reliable value, which was achieved in the quantitative study strand. By means of the qualitative study strand, we were able to characterize low and high grades of complexity. Our theoretical understanding of complexity is based on Perrow’s (1967) framework for the complexity of work in organizations. Accordingly, the extent of complexity depends on the patient’s situation as well as on professional demands to change the situation in accordance with the patients’ aims. Hence, the extent of the complexity of a nursing care situation is to be viewed by taking professionally trained RNs into account. Synthesis of quantitative and qualitative results showed that a smaller portion of patients required high levels of concentrated attention, knowledge, experience and caring from RNs. Nursing managers should support RNs in complex nursing care situations, as these situations can be perceived either as a challenge or as an overwhelming burden (Kentischer et al., 2018). Hence, this instrument may be of use not only to gain monitoring data, but also to red-flag highly complex nursing care situations demanding further attention and support. Future monitoring will reveal whether the number of patients in highly complex nursing care situations will increase.

To our knowledge, the use of an embedded mixed-methods design has not yet been reported in instrument development in the literature. We have chosen this approach where the revaluation of the instrument took place in an advanced stage of its development. Earlier publications about instrument development based on mixed-methods research followed sequential designs to strengthen the instrument’s validity either by defining items with the use of a first qualitative study (Shahbazi Sighaldeh et al., 2019) or by explaining quantitative findings with a second qualitative study (KleinKnecht-Dolf, Spichiger, Müller, Bartholomeyczik, & Spirig, 2017). In contrast, our intention was to comprehend and interpret the instrument’s inter-rater reliability, construct validity, applicability and the meaning of the derived complexity score. The overvalue of the results from the synthesis of the quantitative and qualitative study sections supports this method as useful for future use in a monitoring setting.

4.1 Limitations

However, this study has some limitations. In the quantitative study section, criterion validity could not be investigated because of a lack of instruments in line with this study’s aims. Moreover, insights in the course of complexity over time and an adequate summarized representation of the complexity over the course of hospitalization were unexplored. About the qualitative study section, more cases can be assumed to generate more insights into the perception of complexity of nursing care in
practice. Also, it can be assumed that difficulties in the instrument’s application limit its psychometric properties. Also, this instrument requires some time for reflection, which RNs may lack in day-to-day practice.

5 CONCLUSIONS

In summary, we present a short instrument with evidence of acceptable reliability and validity for the monitoring of patient-related complexity of nursing care in acute care hospitals. However, we found four sub-constructs instead of the originally intended two in the evaluation of the proposed structural and measurement model. This expanded our understanding of patient-related complexity of nursing care based on Perrow’s (1967) framework. Our data highlight that a newly developed instrument should operationalize the theoretical framework in the way that empirical data confirm and advance the theoretical understanding to be studied, which again helps to advance the instrument (Brühl, Planer, & Grebe, 2012).

However, the need to gain familiarity with this instrument shows the necessity of careful instructions and support. Considering the findings of this study may help to purposefully prepare RNs’ applying this new instrument and thus further improve its understanding, applicability and psychometric properties. Additional research is needed to merge the values of the patient-related complexity of nursing care with other monitoring data such as workload, staffing levels, skill-grade mix, nursing outcomes and nursing-sensitive patient outcomes. The results of this study provide a basis for doing so.

ACKNOWLEDGEMENTS

We thank the steering and project members, all the involved nurses and IT technicians from the participating hospitals for their collaboration in the preparation and data collection of this study, Roger Vogt for his personal data management support and Jane White for proofreading the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

EH, MK-D, MM, CK and RS developed the design of the study. MM, EH and MK-D were involved in the collection and management of the data. MM analysed the statistical data, EH performed the qualitative data analysis and the synthesis of the quantitative and qualitative results. EH, MK-D, MM and RS interpreted the results. EH drafted the manuscript. All authors critically revised the manuscript for important intellectual content and approved the final version of the article. EH works in this research programme as a PhD student focussing on "complexity of nursing care" at the University of Witten/Herdecke, Germany. RS and CK are EH’s PhD advisors. RS is the sponsor and MK-D the scientific and project leader of this research programme. MM is the leading statistician.

REFERENCES


Hamada, H., Sekimoto, M., & Imanaka, Y. (2012). Effects of the per


**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Huber E, Kleinknecht-Dolf M, Kugler C, Müller M, Spirig R. Validation of the instrument “Complexity of Nursing Care”—A mixed-methods study. *Nursing Open*. 2019;00:1–13. https://doi.org/10.1002/nop2.383