Kehl, D., Knuth, D., Galea, E., Hulse, L., Sans, J., Valles, L., Seidler, F., Diebe, E., Kecklund, L., Petterson, S., Wolanin, J., Beltowski, G., Preiss, M., Holubova, M., Sofuoglu, T., Baskaya Sofuoglu, Z., Pietrantoni, L., Saccinto, E., & Schmidt, S. (2014). Advancing disaster relief: Development of a self-report questionnaire for firefighters. *International Perspectives in Psychology: Research, Practice, Consultation*, *3(3)*, 167–183. Disponible a: https://doi.org/10.1037/a0036423

**Advancing Disaster Relief: Development of a Self-Report Questionnaire for Firefighters**

DORIS KEHL and DANIELA KNUTH. *Ernst-Moritz-Arndt University of Greifswald*

ED GALEA and LYNN HULSE. *University of Greenwich*

JORDI SANS and LOLA VALLÈS. *Institute of Public Security of Catalonia*

FRANK SEIDLER and EBERHARD DIEBE. *Hamburg Fire and Emergency Service Academy*

LENA KECKLUND and SARA PETTERSON. *MTO Säkerhet AB (MTO Safety AB), Stockholm, Sweden*

JERZY WOLANIN and GRZEGORZ BELTOWSKI. *The Main School of Fire Service (SGSP)*

MAREK PREISS and MARKETA HOLUBOVA. *Prague Psychiatric Centre, Prague, Czech Republic*

TURHAN SOFUOGLU and ZEYNEP BASKAYA SOFUOGLU. *Association of Emergency Ambulance Physicians, Izmir, Turkey*

LUCA PIETRANTONI and ELISA SACCINTO. *University of Bologna*

SILKE SCHMIDT. *Ernst-Moritz-Arndt University of Greifswald*

**Abstract**

During emergencies, firefighters may face challenges such as inappropriate victim responses, poor communication, and lack of needed information. Here we describe the multistep development of a questionnaire for firefighters to identify emergency-related human factors that may impair their professional performance and interfere with procedures during threatening events and evacuations. The resulting self-report questionnaire, titled the Behavior Security Culture–First Responder, comprises a battery of scales plus single questions designed for international use. We report the psychometric properties of the battery of scales and their application across 8 countries: Germany, UK, Spain, Sweden, Poland, Czech Republic, Turkey, and Italy. Exploratory factor analysis and multigroup confirmatory factor analysis were conducted with a total sample of 3,011 firefighters (97% male). Exploratory factor analysis revealed 6 unidimensional scales assessing various aspects of victim behavior, communication with others, and information availability and reliability. Multigroup confirmatory factor analyses, with the 8 countries inserted as groups, confirmed configural and metric invariance, but not scalar invariance. Internal consistency estimates of the 6 scales ranged from α = .77 to .80. Divergent validity was established for all scales in all national samples. The collected data and the questionnaire may be used for developing training programs and to aid in directing resources efficiently.

**Keywords:** emergency evacuation, behavior, firefighter, international, instrument development.

Events that threaten lives and property (e.g., emergencies such as domestic fires, disasters such as floods, and terrorist attacks) present a constant challenge to fire and rescue services. As well as having to navigate their way around the affected property and deal with hazardous conditions such as smoke, flames, water, or debris, firefighters also have to deal with victims' responses to the event. Sometimes victims inside the property and those standing nearby will respond in ways that mitigate the potential harm to themselves and challenges for the emergency

services (e.g., remaining calm and, to the extent possible, removing themselves from the danger in a quick and safe manner). On other occasions, victims may behave in ways that are less constructive.

Effective emergency communication can improve human responses during emergencies and disasters (Kuligowski, 2011), and can prevent or lessen negative outcomes. First, emergency personnel need to be provided with information that allows them to cope efficiently and effectively (Meissner, Luckenbach, Risse, Kirste, & Kirchner, 2002). For victims to respond appropriately to an event, first, their attention needs to be captured and, second, they need to be informed about the type and location of the threat and an appropriate response (Gwynne, 2007; Proulx, 2000). An auditory alarm is often used to capture victims' attention; however, sounds are not recognized universally (Proulx, Laroche, Jaspers-Fayer, & Lavallée, 2001). Different communication channels (i.e., voice communication systems, a person present at the scene, alarm tones or bells, strobe lights, etc.) can be used. However, the application and frequency of communication techniques/channels used during an event could vary among countries (Kuligowski, 2011). Consequently, victims' compliance to warnings and responses to the messages, and thus their evacuation behavior, could differ across nationalities (Kuligowski, 2011). Furthermore, the information must be disseminated by sources the victims trust (Kreps et al., 2005). In addition, firefighters' trust in co-workers, other emergency personnel, and victims to provide them with full and accurate information (Mayer, Davis, & Schoorman, 1995) is especially crucial because firefighters work in risky contexts marked by unpredictability and danger (Colquitt, Lepine, Zapata, & Wild, 2011). A lack of trust in information sources could constitute a stressor and could negatively impact firefighters' work performance (Colquitt et al., 2011).

Thus, encountering responses from victims that are not very constructive, lacking information regarding the situation, and having to deal with communication and trust issues could— either separately or cumulatively— impact firefighters' abilities to carry out their duties. Indeed, a search of the existing literature on firefighters' professional performance and emergency/disaster and evacuation procedures revealed numerous human factors that could affect their performance. The factors included human responses and evacuation behaviors, information and emergency communication at the start of the event and at the scene, firefighters' decisionmaking, victims' willingness to evacuate, victims' compliance and threat assessment, fire education and training, educational modules for professionals in emergency situations, firefighters' self-perceived risk, and emergency personnel's trust in different sources of information. Firefighters could be impacted by such factors in performing their duties and affecting their own health, and the consequences could be both short- and long-term. There could also be secondary effects for the general public's well-being, for example, if fire brigades are short-staffed due to firefighters being on sick leave because of stress.

Although there is a large and growing body of research that has examined human responses to threatening events and evacuations, fewer studies have focused on the firefighters' perspective, or examined the aforementioned issues together. Moreover, few studies have included firefighters from different nations in their samples. Taking all the above into account, a self-administered questionnaire for firefighters, titled the Behavior Security Culture–First Responder (Be-SeCu-FR), was developed as part of the European Union-funded BeSeCu (Behavior–Security–Culture) Project. The questionnaire was designed to investigate human factors that can impede firefighters from carrying out their duties during emergency situations. By examining the frequency of encounters

with such factors, the perceived extent of the problem, and the types of people or agencies involved, it was believed that particularly challenging issues and populations could be identified and prioritized. In turn, the findings could be used to help tailor training programs for emergency service personnel (e.g., within one's own service or multiagency training) and direct resources (e.g., to set up community safety schemes allowing more interaction with targeted sectors of the public). It was important not to simply assume that firefighters' experiences in one country would generalize to experiences in other countries, thus a single instrument was created for international use. Data collected in different countries could help tailor the training and use of resources to local needs. The questionnaire covered a range of topics. Human factors were assessed with a battery of scales and single questions. This article describes the development process of the questionnaire and examines the psychometric properties of the battery of scales that were tested with a sample of more than 3,000 firefighters from eight predominantly European countries.

## Method/Measurement

### *Questionnaire Development*

The BeSeCu-FR was developed via a multistep process, including the aforementioned literature review on firefighter performance and emergency/disaster and evacuation procedures, expert input, focus groups, and pilot testing activities; the latter three steps are described in the following text. Thus, items were generated via a deductive and inductive approach. The study was carried out in Germany, UK, Spain, Sweden, Poland, Czech Republic, Turkey, and Italy between May 2008 and April 2011. Ethical approval was obtained independently in the eight countries from the ethics committees of the data collection centers in the project consortium.

**Expert consultation**. The project consortium consisted of researchers and emergency personnel (e.g., firefighters, emergency medical service providers). Emergency personnel provided information about the knowledge their professions need. The collaboration between researchers and emergency personnel enabled a better distribution of the newly gained knowledge (Fisher, 2010). The project consortium invited other national and international security experts not directly involved in conducting the research to contribute. Those experts supported the study as members of an international advisory group. The advisory group included people from national accident investigation boards and different end-users in the field of evacuation. Consultation with experts was in the form of a structured brainstorming technique.

**Focus groups and interviews**. Focus groups and interviews were conducted across the eight countries with 54 emergency personnel (firefighters, paramedics, physicians, etc.) and key representatives in the context of emergencies/ disasters (i.e., professionals with a good knowledge of civilians' predicaments in these types of event, e.g., counselors, journalists, nongovernmental organizations, politicians, etc.). The interview was divided into three parts. The first part comprised open questions regarding professional background, including experience with large-scale threatening events. The second part comprised a free narrative prompted by the following question "Could you please tell me about your experiences of a typical domestic fire [or situation X] that you were involved in?". The third part was a semi-structured interview with the following domains: setting/characteristics of the situation, recognition that something was happening, decision-making and risk perception, interventions and interactions, and emotional and behavioral reactions.

Demographic and core incident-related information was gathered via questionnaire. Each center taped and then transcribed their interviews/ focus groups. Subsequently, the transcripts were translated into English. From the transcripts, a set of categories was generated as well as domains and items with corresponding answer options derived (Baskaya Sofuoglu, Sofuoglu, Yildirim, & Kehl, 2013).

Pretest. A pilot form of the BeSeCu-FR was developed based on the comprehensive literature review and expert consultations and focus groups/interviews with emergency personnel and key representatives. The pilot form represented 10 domains and comprised 150 items. The pilot form was developed in English and translated into the languages of the participating countries (i.e., Czech, German, Italian, Polish, Spanish, Swedish, and Turkish) using the forward translation method. The questionnaire was pilot-tested simultaneously in eight countries via paper-and-pencil (P&P) questionnaires and constructed for emergency personnel in general (i.e., firefighters, paramedics, emergency physicians, and police officers). A pretest manual was distributed to and followed by all centers. A convenience sample of 260 participants took part in the pretest. On the basis of pretest data, an analysis for missing data and descriptive statistics was conducted. Additionally, qualitative criteria such as acceptance, feasibility, practicability, redundancy, and relevance were considered. The answers to openended items were reviewed to identify commonly occurring topics and to compose new, closed-ended questions. Experts confirmed the applicability of the items to ensure an understanding of the questions in the way they were intended. The percentage of missing items was treated as an appraisal for the acceptance and for the feasibility of items. If the percentage of nonapplicable answers was high, this information was considered as a first hint for the modifying of an item. The decision to keep, revise, or delete a certain item was made via expert consensus, that is, decided unanimously after a discussion during a workshop with representatives from all participating centers. Research results regarding item construction, wording, and questionnaire structure were considered during the process also (Faulmann, Prüfer, & Rexroth, 2009; Handa et al., 2008; Martin et al., 2007). As a result of the pretest, a second draft of the questionnaire was developed.

### *Questionnaire and Battery of Scales*

The second draft was developed in English and afterward translated into the languages of the participating countries using a forwardbackward-forward-translation procedure (Beaton, Bombardier, Guillemin, & Ferraz, 2000). The questionnaire was available in two modes of administration: P&P and online. The modes had identical questions and the layout and response styles were as similar as possible. Questions and layout were identical across all languages. The BeSeCu-FR now comprised 159 items. Those items were split across the following topics or sections: the participant's demographic (e.g., age, gender; 13 items) and work (e.g., working arrangement, rank/role, years of service as a firefighter; 14 items) characteristics; their perceived knowledge and practical experience regarding different events, professional training regarding victim responses and communication with victims, and fire prevention/fire safety education schemes with specific groups (18 items); their risk perception ("on duty" vs. "off duty") concerning different events (12 items); observations of victim behaviour during events (19 items); communication issues (e.g., communication with different emergency services and different sections of the public, plus ways of communicating; 27 items); information issues (e.g., types and sources of information for fires, how information is processed and used; 31 items); their ways of coping with stress and

emotions (9 items); on-duty experiences of being threatened (2 items); and, finally, the psychological impact of the most stressful emergency situation they had professionally attended in the past 10 years (14 items).

Not all of the 159 items were presented using a scale format. For instance, some were presented as single questions with answer formats that produced nominal data (the exact wording of the single questions and their answer options are available from the authors on request). This article focuses primarily on the items that were presented using a scale format and which comprised the following battery of six scales. The first two scales related to the behavior of people at the scene of an emergency, the next two related to communication with people at the scene, and the final two related to the type of information firefighters receive when attending a scene and their trust in different sources of information. Those scales originally consisted of 39 items. Following inspection of the correlations between items and conducting exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) (see *Data Analyses*), the total number of items was reduced to the 20 described in following text.

**Victim Behavior Scale**. The Victim Behavior Scale (VBS) derived from the questionnaire section on observations of victim behaviour and assessed the frequency of observing different nonconstructive victim responses. It was introduced by the following question: “A number of statements that emergency personnel have used to describe victims’ reactions in emergency situations are given below. Please read each reaction and assess how frequently you have seen victims display the following reactions.” A 5-point Likert scale (1 = never, 2 = seldom, 3 = sometimes, 4 = often, and 5 = always) was used with respect to the following four reactions: starting the evacuation too late, not evacuating even though it is required, moving too slowly during the evacuation, and incorrect priorities (e.g., fire fighting instead of evacuation).

**Specific Group Behavior Scale**. The Specific Group Behavior Scale (SGBS) also derived from the section on observations of victim behavior. It assessed whether such victim behaviour was more apparent in certain sectors of the public (identified in the earlier focus groups/interviews as displaying different behaviors or needs) by asking: “How frequently have you encountered incorrect behavior where victims were members of the following groups?” Participants used the same 5-point Likert scale to rate the frequency with respect to three groups: children, elderly, and disabled.

**Efficiency of Communication Scale**. The Efficiency of Communication Scale (ECS) derived from the questionnaire section on communication issues and assessed whether any problems were apparent regarding communication between different emergency services. It was introduced by asking: “During emergency situations, how would you rate the efficiency of the communication of firefighters with other emergency services personnel?” A 5-point Likert scale (1 = never efficient and 5 = always efficient) was used with respect to the four different types of service personnel: firefighter colleagues, police, emergency medical service, and civil protection.

**Specific Group Communication Scale**. Also deriving from the section on communication issues, the Specific Group Communication Scale (SGCS) assessed if problems were apparent regarding communication with different sectors of the public. The scale was introduced by the following question: “How often would you say there are communication difficulties during an operation with the following groups?” A 5-point Likert scale (1 = never and 5 = always) was used with respect to three groups: children, elderly, and disabled.

**Advanced Information Scale**. The Advanced Information Scale (AIS) derived from the questionnaire section concerning information issues. It assessed how much or how little information firefighters would typically be provided with regarding the situation in advance of tackling an emergency with the focus on a particular type of event, that is, a fire in a multistory residential building. It was introduced by the following question: “In general, how often is the following information relayed to you in advance of taking action at such an incident?” The 5-point Likert scale was again used with respect to information about three issues: number of people who need to evacuate, what type of action already taken to rescue victims, and what type of action required to be taken to rescue victims.

**Trust Professional Sources Scale**. The Trust Professional Sources Scale (TPSS) was the second scale to derive from the section on information issues and also related to the scenario of tackling a fire in a multistory residential building. It assessed firefighters’ trust in different sources of information. The TPSS was introduced by asking: “In general, how much do you trust the following sources to provide you with full and accurate information in such an event?” Participants rated (on a 5-point Likert scale: 1 = not at all and 5 = extremely) their trust with respect to three sources: police, emergency medical services, other fire service colleagues.

Although, as mentioned earlier, this article focuses on the battery of scales, a couple of points should be noted about some of the single questions on demographic characteristics, as they are relevant to the following analysis. Regarding the question on income, participants could select from three answer options: their income was less than 70% of the average individual annual net income in their country, their income was from 70% to 150% of the average, or their income was greater than 150% of the average (Grabka & Frick, 2008). The actual values representing the percentages of the published figures for each country’s average individual annual net income (GfK GeoMarketing, 2008) were presented in the questionnaire, for example, participants in the UK saw less than £11,849, £11,849 –£25,391, and more than £25,391. Migrant background was assessed using separate questions about the following characteristics: participant’s country of birth, country of birth of the participant’s mother and father, and participant’s citizenship(s). Participant were categorized as having a migrant background in the following analysis if they answered that they were born in another country to the one in which they currently worked, or their mother or father was born in another country, or they were a citizen of more than one country (Schenk et al., 2006).

### *Field Study Recruitment and Sample*

A field study manual was given to each center, including instructions regarding recruitment of firefighters. This included a requirement of nationwide recruitment. Various possible strategies were recommended. Each center used recruitment strategies that were most effective locally (Knuth, Kehl, Stegemann, & Schmidt, 2013). These included: top-down recruitment via fire brigades (i.e.,

brigades' staff intranets, weekly circulars, e-bulletins, displayed adverts for the project with a link to the online questionnaire, distribution of paper copies of the questionnaire), promotional national and regional radio interviews, adverts in professional journals, posts and links on social network Web sites (e.g., Facebook, first-responder forums), presentations at relevant conferences, and word-of-mouth campaigns via personal contacts in the fire and rescue profession, in academia, and in social circles. The BeSeCu Web site also contained information about the BeSeCu study in all the participating countries' respective languages plus a link to the online questionnaire.

Firefighters could be included if they gave their informed consent, their last operation was no longer than 10 years ago, and were at least 18 years of age. This resulted in a convenience sample of 3,011 participants from eight countries. Demographic and work characteristics are provided in Table 1.

There were differences among national samples with respect to participants' demographic and work characteristics:

$gender\ (male\ vs. female)\ X^2(7) = 22.45,\ p < .01, Cramer's\ V = .09;\ age, F(7, 2989) = 80.41,$ $p < .001, \eta^2 = .16;\ education\ (lowest\ formal\ vs. intermediary\ secondary\ vs. higher\ secondary),$ $X^2(14) = 651.64,\ p < .001, Cramer's\ V = .33;\ income\ (< 70\%\ of\ average\ vs. 70\% \leq x \leq$ $150\%\ of\ average\ vs. > 150\%\ of\ average),\ X^2(14) = 756{,}26,\ p < .001,\ Cramer's\ V =$ $.36;\ relationship\ status\ (yes\ vs. no),\ X^2(7) = 80.47,\ p < .001,\ Cramer's\ V =$ $.16; migrant\ background\ (yes\ vs. no), X^2(7) = 47{,}69,\ p < .001,\ Cramer's\ V =$ $.13; working\ arrangement\ (employed\ vs. honorary\ member), X^2(7) = 1277.61,\ p < .001,$ $Cramer's\ V = .65; rank\ (operational\ vs. supervisory\ operational), X^2(7) = 204.98,\ p < .001,$ $Cramer's\ V = .26;\ and\ years\ of\ service,\ F(7{,}2{,}994) = 30.24,\ p < .001, \eta^2 = .07.$

Of the 3,001 participants, 52,2% filled out the P&P version and 47.8% the online version: Czech Republic: P&P $n$ = 238 (78'8%) versus online $n$ = 64 (21.2%); Germany: P&P $n$ = 124 (17,7%) versus online $n$ = 577 (82,3%); Italy: P&P $n$ = 482 (77,4%) versus online $n$ = 141 (22,6%); Poland: P&P $n$ = 130 (34,7%) versus online $n$ = 245 (65.3%); Spain: P&P $n$ = 119 (73.0%) versus online $n$ = 44 (27.0%); Sweden: P&P $n$ = 9 (5.0%) versus online $n$ = 172 (95.0%); Turkey: P&P $n$ = 467 (99.8%) versus online $n$ = 1 (0.2%); and UK: P&P $n$ = 2 (1.0%) versus online $n$ = 196 (99.0%). An effect of country on administration mode was observed, $X^2(7) = 1447.33,\ p < .001,$ $Cramer's\ V = .69$

Table 1. *Sample Demopraphics and Work Characteristics (N = 3,011)*

| | Field study sample ($N$ = 3,011) | | | EFA sample ($N$ = 1,511) | | CFA sample ($N$ = 1,500) | |
|---|---|---|---|---|---|---|---|
| Sample characteristics | *N* | % | M%[a] | *N* | % | *N* | % |
| Male | 2,915 | 97.0 | 0.2 | 1,467 | 97.3 | 1,448 | 96.7 |
| Education level | | | 0.1 | | | | |
| Lowest formal | 150 | 5.0 | | 78 | 5.2 | 72 | 4.8 |
| Intermediary secondary | 681 | 22.6 | | 334 | 22.1 | 348 | 23.1 |
| Higher secondary | 2,176 | 72.4 | | 1,096 | 72.7 | 1,080 | 72.0 |
| Income | | | 1.1 | | | | |
| <70% of average | 262 | 8.8 | | 129 | 8.6 | 133 | 9.0 |
| 70% ≤ x ≤ 150% of average | 1,427 | 47.9 | | 717 | 48.0 | 710 | 47.9 |
| 150% of average | 1,289 | 43.3 | | 649 | 43.4 | 640 | 43.2 |
| In a relationship | 2,369 | 78.9 | 0.3 | 1,203 | 79.8 | 1,166 | 78.0 |
| Nationality | | | 0.3 | | | | |
| Czech | 285 | 9.5 | | 137 | 9.1 | 148 | 9.9 |
| German | 655 | 21.8 | | 340 | 22.6 | 315 | 21.1 |
| Italian | 601 | 20.0 | | 286 | 19.0 | 315 | 21.1 |
| Polish | 364 | 12.1 | | 193 | 12.8 | 171 | 11.4 |
| Spanish | 159 | 5.3 | | 70 | 4.7 | 89 | 5.9 |
| Swedish | 158 | 5.3 | | 83 | 5.5 | 75 | 5.0 |
| Turkish | 462 | 15.4 | | 230 | 15.3 | 232 | 15.5 |
| British | 180 | 6.0 | | 99 | 6.6 | 81 | 5.4 |
| Migrant background | 137 | 4.6 | | 67 | 4.5 | 70 | 4.7 |
| Working arrangement | | | 0.4 | | | | |
| Employed | 2,446 | 81.6 | | 1,212 | 80.6 | 1,234 | 82.6 |
| Honorary member (unpaid) | 552 | 18.4 | | 292 | 19.4 | 260 | 17.4 |
| Rank/role | | | 0.4 | | | | |
| Operational personnel | 1,867 | 62.3 | | 947 | 63.0 | 920 | 61.5 |
| Supervisory operational personnel | 1,131 | 37.6 | | 556 | 37.0 | 575 | 38.5 |
| | *M* | *SD* | M%[a] | *M* | *SD* | *M* | *SD* |
| Age (in years) | 37.23 | 9.58 | 0.5 | 37.04 | 9.62 | 37.42 | 9.54 |
| Years of service | 13.53 | 8.79 | 0.3 | 13.48 | 8.66 | 13.58 | 8.93 |

*Note*. [a] M% = percentage missing; EFA = exploratory factor analysis; CFA = confirmatory factor analysis.

### *Data Analyses*

There were two aims with the analyses: (1) to establish the dimensions within the 20 items involved and (2) to test their suitability for international use. Thus, participants were randomly assigned to two samples: the first ($n$ = 1,511) to investigate the factor structure of the instrument using EFA and the second ($n$ = 1,500) to test for measurement invariance across countries (and, prior to that, administration modes) via CFA. Comparisons of the two samples revealed no significant differences with respect to the examined demographic and work characteristics (Table 1): $gender, X^2(1) = 1.00, p = .33, Cramer's\ V = .02;\ \ age, t(2{,}995) = 1.08, p = .28,\ r = 0.2;\ \ education, X^2(2) = 0.58, p = .75, Cramer's\ V = .01; income, X^2(2) = 0.11, p = .95, Cramer's\ V = .01; relationship\ status, X^2(1) = 1.52, p = .23, Cramer's\ V = .02; nationality, X^2(8) = 8.63, p = .37, Cramer's\ V = .05; working\ arrangement, X^2(1) = 2.02, p = .16, Cramer's\ V = .03; rank, X^2(1) = 0.69, p = .41, Cramer's\ V = 0.2; and\ years\ of\ service, t(3{,}000) = 0.33, p = .74, r = .01$ . Nor were there any significant differentes with respect to administration mode, $X^2(1) = 2.21, p < .137, Cramer's\ V = .03$. Expectation maximization was used in the following data analysis to manage missing data (for discussion, see Schaferm 1997; Schafer & Olsen, 1998).

The 20-item pool was investigated in the first subsample ($n$ = 1,511) using EFA with principal axis factoring (PAF). PAF is recommended if data are not multivariate or normally distributed (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Before conducting EFA, items that did not correlate with any other items or correlated only with a few were excluded. Additionally, any items that correlated very highly with other items ($r > .8$) were eliminated to avoid extreme multicollinearity and singularity (Field, 2010). Because correlations among the scales were expected and present (see Results section), oblique rotation was used (oblimin with Kaiser normalization). The structure matrix that takes into account the relationships between factors (Field, 2010) is displayed in Table 2. Moreover, EFA using oblique rotation was conducted in each country sample. The number of factors was determined and guided by several criteria: (a) theory (e.g., how many factors were expected? Do the extracted factors make theoretical sense? Do the extracted factors make theoretical sense?), (b) Kaiser-Guttman "eigenvalue greater than 1" rule (Kaiser, 1960), and (c) several solutions/models were explored with different numbers of factors as recommended (Costello & Osborne, 2005). All corrected item-total correlation had to be $\geq .3$ and reliability of each scale (Cronbach's alpha) had to be at least .7 (Field, 2010). The solution that was interpretable and theoretically sensible (Fabriga et al., 1999) was adopted.

For each scale item, descriptive statistics including means, standard deviations, percentage of missing items, skewness, kurtosis, corrected item-total correlation, and Cronbach's alpha (if item deleted) were calculated. Additionally, descriptive statistics of the scales, Cronbach's alpha, and scale intercorrelations (Pearson) were examined. Divergent validity of all scales was tested using a quality of life measure, the WHOQOL 8-item index (Power, 2003; Schmidt, Mühlan, & Power, 2006). This instrument was used as a comparison because, like the BeSeCu-FR scales, it taps into an "individual's perceptions …," that is, their satisfaction or dissatisfaction concerning well-being and interactions with other people and the environment "… in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" (WHOQOL Group, 1994), although it does this purely with regard to the participant's own position in life, whereas the BeSeCu-FR scales ask participants to contemplate other people and their positions. Therefore, it was believed that these different measures may have some commonalities but not enough to be strongly correlated. If the correlation coefficient was smaller than .3 (medium effect), divergent validity was considered to be confirmed.

Multigroup confirmatory factor analyses (MGCFAs) were conducted to test for configural, metric (partial factorial invariance), and scalar (strong factorial invariance) invariance across groups. MGCFA using maximum likelihood estimation with AMOS was performed for each scale separately inserting, first of all, the administration modes as groups. It was desirable to test for measurement invariance across P&P and online survey use because, if scalar invariance was supported, it would be justifiable to pool the data collected with these two methods (Davidov & Depner, 2011). Because an effect of country on administration mode was observed (see *Field Study Recruitment and Sample*), MGCFA was performed for each scale only with one national data sample to take into account this confound. The chosen sample was the Polish firefighter sample ($n = 375$) because the use of the two different modes of administration was less disproportionate in this sample than in the other national firefighter samples (P&P: 34.7% vs. online: 65.3%; see *Field Study Recruitment and Sample*).

MGCFAs were performed a second time for each scale separately, this time inserting the countries as groups. This was to test for measurement invariance across the national samples; in

other words, to see whether the scales were measuring the same thing in the same way in participants from the different countries (Milfont & Fischer, 2010). Because there is an argument for a superior (albeit general) construct—that is, all of the BeSeCu-FR scales measure human factors that impede firefighters from carrying out their duties in emergency situations—MGCFA with countries inserted as groups was also conducted with the 20 items from all six scales combined.

The assessment of measurement invariance included the following three levels of invariance: (a) configural invariance (Model 1), that is, different groups construe the scales in a conceptually similar fashion—this was tested by constraining the factorial structure to be equal across groups; (b) metric invariance (Model 2), that is, scale items are calibrated to their scale's construct in the same way by different groups—this was tested by constraining all factor loadings to be equal across groups; and (c) scalar invariance (Model 3), that is, across different groups, the magnitude of a person's response to a scale (i.e., their obtained score) similarly reflects the magnitude of their response to the underlying construct—this was tested by constraining the items' intercepts to be equal across groups (Milfont & Fischer, 2010).

The following absolute fit indices were used to evaluate overall model fit: (a) the normed $X^2$(i.e., likehood ratio test also called $X^2$ test), and the $X^2$ to degrees of freedom ratio ($X^2$/*df*) (Wheaton, Muthén, Alwin, & Summers, 1977) with a $X^2$/*df* ratio of 3:1 or less indicating a good fit (Carmines & McIver, 1981); (b) the comparative fit index (CFI) with values close to .90 indicating an acceptable fit (Bentler & Bonett, 1980) and close to .95 indicating a good fit (Bentler, 1990); and (c) the root mean square error of approximation (RMSEA) with values $\leq .06$ (Hu & Bentler, 1999; Schreiber, Nora, Stage, Barlow, & King, 2006) or at least $\leq .10$ (Meyers, Gamst, & Guarino, 2006) indicating a good fit.

To assess improvement in fit by comparing a target model with a more constrained nested model (i.e., Model 1 vs. Model 2 and Model 2 vs. Model 3), the following incremental fit indices were used: (a) the difference in $X^2$ between two nested models (Bentler, 1990)— significant results for the $X^2$ difference test indicate that the model with the smaller $X^2$ has a statistically better fit (however, the $X^2$ difference test was used only as indicative of significant improvements because with large samples very minor differences might yield a significant test result); (b) the expected cross-validation index (ECVI, Browne & Cudeck, 1989), with lower ECVI values reflecting the model with the better fit (Brown, 2006); and (c) differences in CFI—minimal differences in these global fit measures between two nested models may support a more restricted model (Cheung & Rensvold, 2002).

Data analyses were carried out using SPSS 18 and AMOS.

Table 2. *Descriptive Item Characteristics of the BeSeCu-FR Scales and Results of EFA (N = 1,511)*

| Items | M%[a] | *M* | *SD* | s[b] | k[c] | Factor loading of the structure matrix[d] 1 | 2 | 3 | 4 | 5 | 6 | $h^2$ | CR[e] | α[f] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VBS: Not evacuating even though it is required | 1.6 | 2.42 | 1.04 | 0.19 | −0.84 | **.78** | .13 | −.12 | −.11 | .05 | −.19 | .61 | .64 | .70 |
| VBS: Starting the evacuation too late | 1.4 | 2.77 | 1.02 | −0.09 | −0.75 | **.78** | .21 | −.12 | −.09 | .09 | −.28 | .60 | .66 | .69 |
| VBS: Moving too slowly | 1.7 | 2.50 | 1.02 | 0.14 | −0.75 | **.68** | .17 | −.12 | −.06 | .04 | −.27 | .46 | .59 | .73 |
| VBS: Incorrect priorities | 1.5 | 2.62 | 1.01 | 0.03 | −0.74 | **.54** | .19 | −.14 | −.07 | .09 | −.27 | .30 | .47 | .79 |
| SGCS: Elderly | 1.1 | 3.05 | 0.84 | −0.13 | −0.23 | .20 | **.78** | −.05 | −.07 | .07 | −.40 | .46 | .63 | .69 |
| SGCS: Disabled | 1.5 | 2.77 | 0.90 | 0.25 | −0.09 | .17 | **.73** | −.03 | .01 | .07 | −.43 | .70 | .62 | .70 |
| SGCS: Children | 0.8 | 2.61 | 0.90 | 0.33 | 0.06 | .18 | **.71** | .02 | .05 | −.03 | −.44 | .56 | .60 | .72 |
| ECS: Emergency medical service | 0.7 | 3.98 | 0.79 | −0.68 | 0.66 | −.18 | −.03 | **.81** | .22 | −.31 | .03 | .66 | .68 | .68 |
| ECS: Police | 0.7 | 3.68 | 0.85 | −0.52 | 0.23 | −.15 | −.08 | **.77** | .13 | −.31 | .12 | .60 | .64 | .69 |
| ECS: Firefighter colleagues | 0.7 | 4.28 | 0.68 | −0.98 | 2.11 | −.09 | −.04 | **.63** | .11 | −.29 | .10 | .40 | .53 | .75 |
| ECS: Civil protection | 1.9 | 3.19 | 1.06 | −0.11 | −0.63 | −.09 | .04 | **.60** | .18 | −.22 | −.01 | .37 | .53 | .78 |
| AIS: What type of action required to be taken | 11.6 | 2.92 | 1.22 | 0.08 | −0.85 | −.07 | .00 | .16 | **.83** | −.25 | −.06 | .69 | .68 | .68 |
| AIS: What type of action already taken | 11.7 | 2.99 | 1.09 | 0.04 | −0.51 | −.05 | .01 | .19 | **.80** | −.24 | −.08 | .64 | .67 | .69 |
| AIS: Number of people who need to evacuate | 11.5 | 3.02 | 1.11 | 0.04 | −0.63 | −.16 | −.04 | .19 | **.65** | −.25 | −.02 | .43 | .57 | .79 |
| TPSS: Emergency medical service | 11.6 | 3.93 | 0.76 | −0.73 | 1.05 | −.12 | −.00 | .35 | .32 | **−.91** | −.02 | .83 | .72 | .54 |
| TPSS: Police | 11.3 | 3.72 | 0.81 | −0.49 | 0.39 | −.12 | −.06 | .34 | .28 | **−.77** | .04 | .59 | .65 | .64 |
| TPSS: Other fire service colleagues | 11.6 | 4.42 | 0.60 | 1.08 | 2.62 | −.01 | −.03 | .20 | .14 | **−.52** | .02 | .28 | .47 | .82 |
| SGBS: Disabled | 2.9 | 2.09 | 1.00 | 0.67 | −0.18 | .26 | .43 | −.05 | .11 | .03 | **−.83** | .53 | .68 | .67 |
| SGBS: Children | 1.7 | 2.35 | 1.08 | 0.37 | −0.79 | .31 | .44 | −.04 | .05 | .00 | **−.74** | .51 | .63 | .72 |
| SGBS: Elderly | 2.3 | 2.82 | 0.99 | −0.15 | −0.64 | .26 | .43 | −.11 | −.01 | .06 | **−.67** | .62 | .60 | .75 |
| Initial eigenvalue | | | | | | 3.96 | 3.23 | 1.90 | 1.84 | 1.40 | 1.09 | | | |
| Percentage of variance[g] | | | | | | 17.6 | 14.0 | 7.2 | 7.0 | 5.0 | 3.4 | | | |
| | | | | | | VBS | SGCS | ECS | AIS | TPSS | SGBS | | | |

*Note.* Factor 1 = Victim Behavior Scale (VBS); Factor 2 = Specific Group Communication Scale (SGCS); Factor 3 = Efficiency of Communication Scale (ECS); Factor 4 = Advanced Information Scale (AIS); Factor 5 = Trust Professional Sources Scale (TPSS); Factor 6 = Specific Group Behavior Scale (SGBS); $h^2$ = communalities after extraction.
[a] M% = percentage missing. [b] s = skewness. [c] k = kurtosis. [d] Items are ordered according to their factor loadings from highest to lowest and grouped according to factor. Boldface indicates to which factor the factor loadings belong. [e] CR = corrected item-total correlation with respect to the corresponding factor. [f] α = Cronbach's alpha (if item deleted) with respect to the corresponding factor. [g] When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

## Results

### *EFA Outcomes*

EFA with PAF using oblique rotation in the total EFA sample indicated a six-factor solution (Table 2). The six factors were named as follows: Victim Behavior Scale (VBS), Specific Group Behavior Scale (SGBS), Efficiency of Communication Scale (ECS), Specific Group Communication Scale (SGCS), Advanced Information Scale (AIS), and Trust Professional Sources Scale (TPSS). EFA with oblique rotation in all national samples replicated the scales VBS, ECS, AIS, and TPSS. Differences were revealed with respect to the items on observations of incorrect behavior and communication involving different sectors of the public. In most national samples, these items

loaded on two factors, as they did in the entire EFA sample, named as SGBS and SGCS. In the Polish sample, however, these items loaded on one factor.

From this, we concluded that six scales were present across all countries, namely, VBS, SGBS, ECS, SGCS, AIS, and TPSS, although with the SGBS and SGCS suffering from minor irregularities. For each of the 20 items making up the six scales, the factor loadings, Cronbach's alpha (if item deleted), and corrected item-total correlations are presented in Table 2.

To demonstrate the unidimensionality of each scale, EFA with PAF using oblique rotation was conducted for each scale separately. For all six scales, a one-factor solution was revealed: 48.72% of the total variance was explained in the VBS (eigenvalue = 2.43), 56.39% in the SGBS (eigenvalue = 2.12), 49.83% in the ECS (eigenvalue = 2.47), 54.26% in the SGCS (eigenvalue = 2.08), 57.67% in the AIS (eigenvalue = 2.14), and 51.55% in the TPSS (eigenvalue = 1.88). Descriptive statistics and internal consistency values for each scale are provided in Table 3. Cronbach's alpha ranged from .77 to .80 in the EFA sample and from .75 to .81 in the CFA sample. In each national sample, Cronbach's alpha for the scales was at least .7. One exception was the SGBS in the UK sample.

### *MGCFA Outcomes*

When measurement invariance was tested with the two administration modes as groups, the global fit measures were good and suggested that the configural, metric, and scalar invariance models were supported by the data (Table 4). Consequently, it is justifiable to pool the data collected with these two methods and compare group means.

When measurement invariance was tested for each scale with the eight countries inserted as groups, fit indices revealed that configural invariance and metric invariance were supported, but scalar invariance was not (Table 5). Consequently, this allows comparisons of correlates (covariances and regression coefficients) across national groups within each scale but not comparisons of means across these national groups.

When measurement invariance was tested with the 20 items of the six scales combined and having the countries as groups, the global fit measures were acceptable and suggested that the configural invariance model ($X^2(1{,}249) = 2{,}184.47, p < .001, X^2ldf = 1.75, CFI = .91, RMSEA = .022\ [90\%\ CI: .021 - .024], ECVI = 2.26$) and the metric invariance model ($X^2(1{,}347) = 2{,}441.60, p < .001, X^2ldf = 1.81, CFI = .90, RMSEA = .023\ [90\%\ CI: .022 - 025], \Delta CFI = .015, ECVI = 2.29$) were supported by the data and should not be rejected. However, the global fit measures suggested that the scalar invariance model should be rejected ($X^2(1{,}487) = 4{,}448.23, p < .001, X^2ldf = 2.99, CFI = .72, RMSEA = .037\ [90\%\ CI: .035 - .038], \Delta CFI = .180, ECVI = 3.45$).

Table 3. *Descriptive Scale Characteristics in EFA Subsample (N = 1,511)*

| Scale | M%[a] | *M* | *SD* | s[b] | k[c] | Items | α | α in national samples |
|---|---|---|---|---|---|---|---|---|
| VBS | 3.0 | 2.58 | 0.79 | −.03 | −.43 | 4 | .78 | .67, .71–.85 |
| SGBS | 3.3 | 2.42 | 0.86 | .33 | −.29 | 3 | .79 | .54, .73–.85 |
| ECS | 2.7 | 3.78 | 0.66 | −.25 | .33 | 4 | .78 | .64, .71–.83 |
| SGCS | 1.9 | 2.81 | 0.74 | .18 | .46 | 3 | .78 | .61, .66, .71–.85 |
| AIS | 12.0 | 2.98 | 0.96 | .10 | −.44 | 3 | .80 | .63, .65, .73–.85 |
| TPSS | 12.2 | 4.02 | 0.61 | −.48 | .63 | 3 | .77 | .68, .74–.86 |

[a] M% = percentage missing. [b] s = skewness. [c] k = kurtosis, α in CFA subsample ($N$ = 1,500): .79 (VBS), .76 (SGBS), .81 (AIS), .75 (ECS), .78 (SGCS), .80 (TPSS).

Table 4. *Fit Indices for Tests of Measurement Invariance Across the Two Administration Modes in the Polish Firefighter Sample (N = 375)*

| Scale | Comparison | $\chi^2$ (*df*) | $\chi^2/df$ | Δ $\chi^2$ (Δ*df*) | CFI (Δ CFI) | RMSEA (90% CI) | ECVI | Decision |
|---|---|---|---|---|---|---|---|---|
| VBS | | | | | | | | |
| Model 1: Configural | — | 17.33** (4) | 4.33 | — | .969 (–) | .945 (.052–.142) | .175 | Accept |
| Model 2: Metric | M1 versus M2 | 18.31* (7) | 2.62 | +0.99[n.s.] (3) | .973 (.004) | .066 (.030–.104) | .162 | Accept |
| Model 3: Scalar | M2 versus M3 | 23.44* (11) | 2.13 | +5.13[n.s.] (4) | .971 (.002) | .055 (.023–.086) | .154 | Accept |
| SGBS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 5.68[n.s.] (2) | 2.84 | +5.68[n.s.] (2) | .990 (.001) | .070 (.000–.141) | .101 | Accept |
| Model 3: Scalar | M2 versus M3 | 10.92[n.s.] (5) | 2.18 | +5.24[n.s.] (3) | .983 (.006) | .056 (.000–.102) | .099 | Accept |
| ECS | | | | | | | | |
| Model 1: Configural | — | 20.97*** (4) | 5.24 | — | .961 (–) | .093 (.056–.134) | .141 | Accept |
| Model 2: Metric | M1 versus M2 | 20.97** (7) | 3.00 | +0.00[n.s.] (3) | .968 (.007) | .064 (.034–.096) | .129 | Accept |
| Model 3: Scalar | M2 versus M3 | 20.97* (11) | 1.91 | +.000[n.s.] (4) | .977 (.009) | .043 (.012–.071) | .113 | Accept |
| SGCS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 1.11[n.s.] (2) | 0.55 | +1.11[n.s.] (2) | 1.000 (.000) | .000 (.000–.086) | .089 | Accept |
| Model 3: Scalar | M2 versus M3 | 2.05[n.s.] (5) | 0.41 | +0.94[n.s.] (3) | 1.000 (.000) | .000 (.000–.041) | .075 | Accept |
| AIS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 1.77[n.s.] (2) | 0.88 | 1.77[n.s.] (2) | 1.000 (.000) | .000 (.000–.099) | .091 | Accept |
| Model 3: Scalar | M2 versus M3 | 13.76* (5) | 2.75 | 11.99** (3) | .970 (.030) | .069 (.026–.113) | .107 | Accept |
| TPSS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 0.75[n.s.] (2) | 0.37 | 0.75[n.s.] (2) | 1.000 (.000) | .000 (.000–.077) | .088 | Accept |
| Model 3: Scalar | M2 versus M3 | 13.52* (5) | 2.70 | 12.77** (3) | 0.986 (.014) | .068 (.025–.112) | .106 | Accept |

*Note.* VBS = Victim Behavior Scale; SGBS = Specific Group Behavior Scale; ECS = Efficiency of Communication Scale; SGCS = Specific Group Communication Scale; AIS = Advanced Information Scale; TPSS = Trust Professional Sources Scale; n.s. = not significant.
[a] Minimum was achieved, i.e., the model was fitted successfully—probability level cannot be computed.
* $p < .05$. ** $p < .01$. *** $p < .001$.

### *Divergent Validity*

The intercorrelations (Pearson) of the scales with the WHOQOL 8-item index are displayed in Table 6. All scales either did not correlate or only had a small correlation (regardless of significance) with the WHOQOL 8-item index; this was also the case in all national samples.

## Discussion

The BeSeCu-FR, an international questionnaire for firefighters, was developed to evaluate factors that may impact firefighters' professional performance during emergencies and evacuations. A strength of the questionnaire is that it was developed simultaneously across several nations, a process that is all too rare in instrument development.

The aim of this article was to describe the development of the BeSeCu-FR and to examine the psychometric properties of the battery of scales, tested in eight countries. The results supported six scales assessing: observations of victims displaying incorrect behavior, the incidence of such behavior in different sectors of the public, the efficiency of communication with other emergency services personnel, communication difficulties with different sectors of the public, information

relayed to firefighters in advance of starting to tackle an incident, and trust in professional sources of information. Considering a Cronbach's alpha of at least .7 as acceptable (Gliem & Gliem, 2003), internal consistency values of all scales were satisfactory.

MGCFA suggested that the P&P and online versions of the questionnaire were measuring the same thing in the same way, meaning that pooling the data collected with these two different methods and conducting comparisons of scale means across these modes are justified (Davidov & Depner, 2011). It might be valuable if future studies try to replicate those findings and include additional techniques of inquiry such as telephone or personal interviews (Davidov & Depner, 2011). If such additional techniques were also found to be equivalent measurement methods, then this would mean that more diverse recruitment strategies can be applied.

MGCFA also suggested a certain level of measurement invariance when each scale was examined with the eight countries included as the groups. Fit indices confirmed that configural invariance was supported for each scale, indicating that the factorial structures of the constructs were equal across the national groups. As configural invariance was supported, the factor loadings were then constrained to be equal to test for metric invariance. The metric invariance model of each scale had good fit indices (e.g.,$X^2 ldf \leq 3; RMSEA < .06; CFI > .95$), but the $X^2$ tests were significant, indicating that the imposition of constraints (equal factor loadings across national groups) resulted in statistically significant decreases in the fit of Model 2 compared with Model 1. However, as stated earlier,the $X^2$ difference test has limitations (see *Data Analyses*). Taking the other comparative fit indices (e.g., $\Delta CFI$) into account, the results indicated overall that the factor loadings were the same across national groups within each scale. Consequently, this allows comparisons of correlates (covariances and regression coefficients) across national groups within each scale. The next step taken was to test for scalar invariance for each scale. The overall goodness-of-fit indices and the tests of differences in fit between Models 2 and 3 did not support scalar invariance across the different national groups. Scalar invariance was similarly rejected when all 20 items of the six scales were included together in the analysis. These results suggest that latent means cannot be meaningfully compared across the eight countries. However, it may still be possible to compare the scale means across a smaller set of countries. Whether scalar invariance may hold for some or all scales in a smaller set of countries will be investigated in future studies. Nevertheless, it is worth considering here that achieving strong (scalar) measurement invariance might not be realistic in the circumstances examined with the BeSeCu-FR. In some cultures, it might be acceptable for professionals to be critical of colleagues and/or the people they serve. In other cultures, it might not be acceptable to express dissatisfaction even if it is felt privately. Thus, certain biases, for example, those related to social desirability, might have an impact on responses at the item level for one or more national groups (Bollen, 1989; Gregorich, 2006). In future research, it might be advisable to test social desirability and, if biases were detected, control for their influence when assessing measurement invariance.

Table 5. *Fit Indices for Tests of Measurement Invariance Across Eight National Groups (N = 1,500)*

| Scale | Comparison | $\chi^2$ (*df*) | $\chi^2/df$ | $\Delta\chi^2$ ($\Delta df$) | CFI (ΔCFI) | RMSEA (90%CI) | ECVI | Decision |
|---|---|---|---|---|---|---|---|---|
| VBS | | | | | | | | |
| Model 1: Configural | — | 21.77$^{n.s.}$ (16) | 1.36 | — | .997 (–) | .016 (.000–.030) | .143 | Accept |
| Model 2: Metric | M1 versus M2 | 75.21*** (37) | 2.03 | +53.43*** (21) | .978 (.019) | .026 (.018–.035) | .151 | Accept |
| Model 3: Scalar | M2 versus M3 | 286.26*** (64) | 4.47 | +211.06*** (27) | .879 (.108) | .048 (.043–.054) | .256 | Reject |
| SGBS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 44.40*** (14) | 3.17 | +44.40*** (14) | .973 (.027) | .038 (.026–.051) | .108 | Accept |
| Model 3: Scalar | M2 versus M3 | 415.11*** (35) | 11.86 | +370.71*** (21) | .661 (.312) | .085 (.078–.093) | .328 | Reject |
| ECS | | | | | | | | |
| Model 1: Configural | — | 20.12$^{n.s.}$ (16) | 1.25 | — | .997 (–) | .013 (.000–.029) | .142 | Accept |
| Model 2: Metric | M1 versus M2 | 62.55** (37) | 1.69 | +42.43** (21) | .983 (.014) | .022 (.012–.031) | .142 | Accept |
| Model 3: Scalar | M2 versus M3 | 458.85*** (65) | 7.06 | +396.30*** (28) | .738 (.245) | .064 (.058–.069) | .371 | Reject |
| SGCS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 42.63*** (14) | 3.05 | +42.63*** (14) | .976 (.024) | .037 (.025–.050) | .106 | Accept |
| Model 3: Scalar | M2 versus M3 | 198.78*** (35) | 5.68 | +156.15*** (21) | .862 (.114) | .056 (.049–.064) | .183 | Reject |
| AIS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 42.33*** (14) | 3.02 | +42.33*** (14) | .981 (.019) | .037 (.024–.050) | .106 | Accept |
| Model 3: Scalar | M2 versus M3 | 289.93*** (35) | 8.28 | +247.60*** (21) | .826 (.155) | .070 (.063–.077) | .244 | Reject |
| TPSS | | | | | | | | |
| Model 1: Configural | — | 0.00[a] (0) | — | — | 1.000 (–) | — | .097 | Accept |
| Model 2: Metric | M1 versus M2 | 45.36*** (14) | 3.24 | +45.36*** (14) | .983 (.017) | .039 (.027–.052) | .108 | Accept |
| Model 3: Scalar | M2 versus M3 | 246.03*** (35) | 7.03 | +200.67*** (21) | .888 (.095) | .064 (.056–.071) | .215 | Reject |

*Note.* VBS = Victim Behavior Scale; SGBS = Specific Group Behavior Scale; ECS = Efficiency of Communication Scale; SGCS = Specific Group Communication Scale; AIS = Advanced Information Scale; TPSS = Trust Professional Sources Scale; n.s. = not significant.
[a] Minimum was achieved, i.e., the model was fitted successfully—probability level cannot be computed.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6. *Zero-Order Correlations (Pearson) of Scale Means (N = 1,511)*

| Scale | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. VBS | — | | | | | | |
| 2. SGBS | .307*** | — | | | | | |
| 3. ECS | −.147*** | −.064* | — | | | | |
| 4. SGCS | .200*** | .465*** | −.017 | — | | | |
| 5. AIS | −.098*** | .048 | .191*** | .005 | — | | |
| 6. TPSS | −.097*** | −.027 | .328*** | −.032 | .283*** | | |
| 7. WHOQOL8 | −.026 | −.007 | .156*** | −.016 | .013 | .109*** | — |

*Note.* VBS = Victim Behavior Scale; SGBS = Specific Group Behavior Scale; ECS = Efficiency of Communication Scale; SGCS = Specific Group Communication Scale; AIS = Advanced Information Scale; TPSS = Trust Professional Sources Scale; WHOQOL8 = WHOQOL 8-item index.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Divergent validity was confirmed in all national samples, because all scales either did not correlate or only had a small correlation with the WHOQOL 8-item index. Because the questionnaire length was kept as short as possible by limiting the use of additional measures to have only content directly related to the study's purpose (Worthington & Whittaker, 2006), information on convergent validity is not available.

Of interest, the present data revealed a positive relationship between the SGCS and the SGBS. Thus, communication difficulties between firefighters and certain types of victims during an operation were associated with reactions displayed by certain types of victims that did not seem correct given the circumstances (e.g., starting evacuation too late). Although the present data do not allow inferences about causality, this does suggest that firefighters could improve their security-related communications and instructions with victims on site and that the communications and instructions should be tailored to different sectors of the public. Also, if firefighters have problems communicating with colleagues and professionals in other agencies (ECS), this could indicate a need for better training in communication. Communication issues could, for example, be due to

personal shortcomings in social skills, which could be eliminated through training. Also the use of unclear or ambiguous terminology could be eliminated through training. Furthermore, the present data revealed a positive relationship between the ECS and the TPSS. Thus, the efficiency of communication toward other emergency service personnel may be related to the trust toward them. The present study focused on the questionnaire's content and psychometric properties, and content-based questions, as well as differences between the nationalities, will be analyzed and published in subsequent papers.

Also, in future studies, we will model through structural equation modeling the ways in which the scales might be related to each other.

Finally, we note some limitations of the present study. First, the sample is a convenience sample and differences between national samples with respect to participants' demographic and work characteristics were present. Such differences, particularly differences in age, could have influenced our questionnaire development. Second, gender-related issues could not be considered, because only 3% of the firefighter sample was female. However, this represents the usual percentage of women among firefighters. For example, in the UK, nearly 3.9% of firefighters are women (Department for Communities and Local Government, 2011), and 3.7% in the United States (Hulett, Bendick, Thomas, & Moccio, 2008). Third, pretesting was conducted with emergency personnel in general, whereas the field study questionnaire was developed for firefighters only. This was a consequence of the development process, because not all generated items were applicable for all groups of investigated first responders. As the main focus was on firefighters, the field study questionnaire was made specifically for them. Fourth, dividing the original sample randomly to investigate the scales with EFA in one and confirm them with CFA in the other subsample may have increased the probability that the scales functioned in the same way; thus, acquiring a completely new sample would be a more stringent test of the psychometric properties. Moreover, further procedures than those conducted here (e.g., crossvalidation) may arguably be required. Fifth, participants who filled out the P&P version had more missing values ($M = 3.04, SD = 6.25$) than participants who filled out the online version ($M = .80, SD = 1.91$). However, these results are in accordance with the existing literature on questionnaires and response rates (Denscombe, 2006; Kongsved, Basnov, Holm- Christensen, & Hjollund, 2007).

### *Perspectives*

The collected data and the battery of scales may be used for developing training programs and to aid in directing resources efficiently. The firefighters' performance on the battery of scales as a whole may enable an assessment of how much firefighters are impeded in their duties. Furthermore, to assess how problematic the different types of challenges are, so that training and educational programs can be tailored accordingly, the scales can be used one by one. So, for example, if emergency services believe that certain areas are especially problematic and want to assess whether their training/educational programs have any success, it might be practicable to use just one or two scales from the whole BeSeCu-FR. As well as being of benefit to the fire and rescue services, we believe that the findings could better inform other professionals involved in managing emergency evacuation situations. As a consequence, multiple agencies would be in a position to devise strategies to help staff cope with issues that impact their performance and, in turn, people's safety and well-being could be enhanced.

**References**

Baskaya Sofuoglu, Z., Sofuoglu, T., Yildirim, E., & Kehl, D. (2013). Human behaviour and culture in emergency situations: Focus group analysis of first responders and key representatives. In S. Schmidt, & E. Galea (Eds.), *Behaviour-security-culture. Human behaviour in emergencies and disasters: A cross-cultural investigation*. Lengerich, Germany: Papst Science Publishers.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*, 25, 3186 –3191. doi: 10.1097/00007632-200012150-00014

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588– 606. doi:10.1037/0033-2909.88.3.588

Bollen, K. A. (1989). *Structural equations with latent variables*. Oxford, England: Wiley.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, 24, 445–455. doi:10.1207/s15327906mbr2404_4

Carmines, E. G., & McIver, J. D. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohinstedt & E. F. Borgotta (Eds.), *Social measurements: Current issues* (pp. 65–115). Beverly Hills, CA: Sage.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluation of goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255. doi:10.1207/S15328007sem0902_5

Colquitt, J. A., Lepine, J. A., Zapata, C. P., & Wild, R. E. (2011). Trust in typical and high- reliability contexts: Building and reacting to trust among firefighters. *Academy of Management Journal*, 54, 999–1015. doi:10.5465/amj.2006.0241

Costello, A., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10, 1–9.

Davidov, E., & Depner, F. (2011). Testing for measurement equivalence of human values across online and paper-and-pencil surveys. *Quality and Quantity*, 45, 375–390. doi:10.1007/s11135-009-9297-9

Denscombe, M. (2006). Web-based questionnaires and the mode effect. An evaluation based on completion rates and data contents of near-identical questionnaires delivered in different modes. *Social Science Computer Review*, 24, 246 –254. doi: 10.1177/0894439305284522

Department for Communities and Local Government. (2011). Retrieved from http://www.communities.gov.uk/newsroom/factsandfigures/communitiesfire1/facts/fire/?id=1951360. Retrieved April 11, 2011.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272–299. doi: 10.1037/1082-989X.4.3.272

Faulmann, F., Prüfer, P., & Rexroth, M. (2009). *Was ist eine gute Frage? Die systematische evaluation der fragenqualität* (1 ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Field, A. (2010). *Discovering statistics using SPSS* (3 ed.). Los Angeles, CA: SAGE Publications.

Fisher, H. W. (2010). Future research needs. In *Response to disaster: Fact versus fiction and its perpetuation. The sociology of disaster* (pp. 201–202). Lanham, MD: University Press of America.

GfK GeoMarketing. (2008). Europa bleibt gespalten. GfK Kaufkraft für 41 europäische Länder. Retrieved from http://www.gfk.com/imperia/md/content/presse/pm_kaufkraft_europa_2008_dfin.pdf. Retrieved August 3, 2010.

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's Alpha reliability coefficient for likert-type scales*. Paper presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, Columbus, OH.

Grabka, M. M., & Frick, J. R. (2008). Schrumpfende Mittelschicht – Anzeichen einer dauerhaften Polarisierung der verfügbaren Einkommen? *Deutsches Institut für Wirtschaftsforschung Berlin - Wirtschaft Politik Wissenschaft, 10,* 101–108.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44, S78–S94. doi: 10.1097/01.mlr.0000245454.12228.8f

Gwynne, S. M. V. (2007). *Optimizing fire alarm notification for high risk groups (Summary Report)*. Quincy, MA: The Fire Protection Research Foundation.

Handa, V. L., Barber, M. D., Young, S. B., Aronson, M. P., Morse, A., & Cundiff, G. W. (2008). Paper versus web-based administration of the Pelvic Floor Distress Inventory 20 and Pelvic Floor Impact Questionnaire 7. *International Urogynecology Journal and Pelvic Floor Dysfunction*, 19, 1331–1335. doi:10.1007/s00192-008-0651-6

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. doi:10.1080/10705519909540118

Hulett, D. M., Bendick, M., Thomas, S. Y., & Moccio, F. (2008). A national report card on woman in firefighting. International Association of Woman in Fire and Emergency Services. Retrieved from http://www.i-women.org/images/pdf-files/35827WSP.pdf. Retrieved March 10, 2014.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measuremen*t, 20, 141–151. doi: 10.1177/001316446002000116

Knuth, D., Kehl, D., Stegemann, K., & Schmidt, S. (2013). Recruitment strategies across all centers. In S. Schmidt & E. Galea (Eds.), *Behaviour-Security-Culture. Human behaviour in emergencies and disasters: A cross-cultural investigation*. Lengerich, Germany: Papst Science Publishers.

Kongsved, S. M., Basnov, M., Holm-Christensen, K., & Hjollund, N. H. (2007). Response rate and completeness of questionnaires: a randomized study of internet versus paper-and-pencil versions. *Journal of Medical Internet Research*, 9, e25. doi:10.2196/jmir.9.3.e25

Kreps, G. L., Alibek, K., Bailey, C., Neuhauser, L., Rowan, K. E., & Sparks, L. (2005). The critical role of communication to prepare for biological threats: prevention, mobilization, and response. In H. D. O'Hair, R. L. Heath, & G. R. Ledlow (Eds.), *Community preparedness and response to terrorism. v. 3. Communication and the Media*. Westport, CT: Praeger.

Kuligowski, E. (2011). Communicating the Emergency: Preliminary findings on the elements of an effective public warning message. In W. Jaskó[swsl]lowski & P. Kêpka (Eds.), *Emergency evacuation of people from buildings*. Warszawa, Poland: BEL Studio Sp. z o.o.

Martin, E., Hunter Childs, J., DeMaio, T., Hill, J., Reiser, C., Gerber, E., . . . Dillman, D. (2007). *Guidelines for designing questionnaires for administration in different modes*. Washington, DC: U.S. Census Bureau.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20, 709–734. doi:10.5465/amr.1995.9508080335

Meissner, A., Luckenbach, T., Risse, T., Kirste, T., & Kirchner, H. (2002). *Design challenges for an integrated disaster management communication and information system*. First IEEE Workshop on Disaster Recovery Networks (DIREN). Retrieved from http://www.comet.columbia.edu/diren/

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation (Vol. 45)*. Thousand Oaks, CA: Sage.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3, 111–121.

Power, M. (2003). Development of a common instrument for quality of life. In A. Nosiko & C. Gudex (Eds.), *EUROHIS. Developing common instruments for health surveys* (pp. 145–161). Amsterdam: Published on behalf of the World Health Organization Regional Office for Europe by IOS Press.

Proulx, G. (2000). *Strategies for ensuring appropriate occupant response to fire alarm signals (Construction Technology Update No. 43)*. Ottawa, Canada: Institute for Research in Construction, National Research Council Canada.

Proulx, G., Laroche, C., Jaspers-Fayer, F., & Lavallée, R. (2001). Fire Alarm Signal Recognition (Internal Report No. 828). Ottawa, Canada: Institute for Research in Construction, National Research Council Canada.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall. doi:10.1201/9781439821862

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33, 545–571. doi:10.1207/s15327906mbr3304_5

Schenk, L., Bau, A. M., Borde, T., Butler, J., Lampert, T., Neuhauser, H., . . . Weilandt, C. (2006). Mindestindikatorensatz zur Erfassung des Migrationsstatus. Empfehlungen für die epidemiologische Praxis. *Bundesgesundheitsblatt – Gesundheitsforschung - Gesundheitsschutz, 49*, 853–860. doi:10.1007/s00103-006-0018-4

Schmidt, S., Mühlan, H., & Power, M. (2006). The EUROHIS-QOL 8-item index: Psychometric results of a cross-cultural field study. *European Journal of Public Health*, 16, 420 – 428. doi:10.1093/eurpub/cki155

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*, 323–338. doi:10.3200/joer.99.6.323-338

Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing the reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology* (pp. 84–136). San Francisco, CA: Jossey-Bass. doi:10.2307/270754

WHOQOL Group. (1994). Development of the WHOQOL: Rationale and current status. *International Journal of Mental Health*, 23, 24–56.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research. *The Counseling Psychologist*, 34, 806 – 838. doi:10.1177/0011000006288127