Article

Feasibility, Interrater Reliability and Internal Consistency of the German Environmental Audit Tool (G-EAT)

Anne Fahsold [1,2,*], Kathrin Schmüdderich [1,2], Hilde Verbeek [3], Bernhard Holle [1,2] and Rebecca Palm [2]

1   Deutsches Zentrum für Neurodegenerative Erkrankungen e.V. Standort Witten, 58453 Witten, Germany; kathrin.schmuedderich@dzne.de (K.S.); bernhard.holle@dzne.de (B.H.)
2   Department of Nursing Science, Faculty of Health, Witten/Herdecke University, 58455 Witten, Germany; rebecca.palm@uni-wh.de
3   Department of Health Services Research, Maastricht University, 6229 GT Maastricht, The Netherlands; h.verbeek@maastrichtuniversity.nl
*   Correspondence: anne.fahsold@dzne.de

Abstract: Dementia-specific environmental design has the potential to positively influence capabilities for daily living and quality of life in people with dementia living in nursing homes. To date, no reliable instrument exists for systematically assessing the adequacy of these built environments in Germany. This study aimed to test the adapted version of the Environmental Audit Tool—High Care (EAT-HC)—the German Environmental Audit Tool (G-EAT)—with regard to its feasibility, interrater reliability and internal consistency. The G-EAT was applied as a paper-pencil version in the German setting; intraclass correlation coefficients at the subscale level ranged from 0.662 (III) to 0.869 (IV), and 42% of the items showed at least substantial agreement (Cohen's kappa $\geq$ 0.60). The results indicate the need to develop supplementary material in a manual that illustrates the meaning of the items and practical implications regarding dementia-specific environmental design. Furthermore, the intersectionality of built and physical environments must be considered when interpreting G-EAT results in future research and applications to residential long-term care practice.

Keywords: dementia-specific environment; environmental design; assessment instrument; long-term care; dementia; reliability; feasibility

1. Introduction

Although 80.2% of the 4,127,605 people in need of care in Germany receive support at home from relatives (64.0%) and outpatient services (29.7%) [1], residential long-term care is an irreplaceable pillar of the German care system. A total of 14,494 nursing homes provided 877,162 full-time long-term care places in 2019 [2]. Hoffmann et al. (2014) identified that 51.8% (95%-CI 50.4–53.3) [3] of nursing home residents are diagnosed with dementia. Beyond this, the number of residents living with mild cognitive impairments or undiagnosed dementia cannot be quantified, but it is estimated to be much higher. Residential long-term care facilities in Germany differ in various respects, including in terms of their size, the number of full-time long-term care places offered and their living concepts. For example, the majority of nursing homes provide care in integrative living units where people with and without dementia live together. Only approximately 30% of living units exclusively accommodate people with dementia and are designated dementia special care units [4]. Nursing homes in Germany are often organized in living units—most with 2–3 per facility—but their characteristics are not clearly defined and vary across facilities.

The built environment has been known for several years to be a key element of dementia-specific care [5]. If nursing homes are designed with people with dementia in mind, this may help residents maintain their independence in daily activities and improve their quality of life [5]. Specifically, the advantages of such an adapted environment have

been addressed in various literature reviews in recent years [5–8]. Several factors, such as special layout and oriental cues or a homelike environment, can influence this support in a positive way. These environmental factors have been examined in a number of studies and have led to the development of various design guidelines and assessment instruments [9–11]. To describe dementia-specific environments in detail, systematic assessment tools are needed to map the different aspects of dementia-specific design. Thus far, no instrument of this kind exists in Germany. In reviewing available tested and established English language instruments, the Environmental Audit Tool—High Care (EAT-HC) [12] showed the most similarities to the target setting regarding understanding of care of people with dementia [13].

The EAT-HC is the enhanced version of Fleming and Bennett's Environmental Audit Tool (EAT). In 2011, the creators developed the EAT based on a literature review [8,14]. The ten subscales of the instrument constitute the so-called "Key Design Principles": (1) risk reduction, (2) human scale, (3) seen and be seen, (4 and 5) positive and negative stimulation levels, (6) movement and engagement, (7) familiarity, (8 and 9) variety of places in and around the unit, and (10) design and way of life [12] (pp. 198–200). The subscales contain between 2 and 19 items for which a subscore is calculated. All items, with the exception of those of Key Design Principle "Design and way of life", form the unweighted overall score of the EAT-HC. The items "What is the vision/purpose of the unit for people with dementia?" and "How well does the built environment enable this to happen?" [12] (p. 264) serve to explain a health care team's goals for the particular living unit to be assessed. The results of the instrument reflect these goals, and the need for action (e.g., rebuilding of outside areas or improvement of wayfinding markers) may be deduced accordingly. For the EAT-HC, another 56 items were formulated [14–16]. In testing interrater reliability, 50 items with an interrater agreement of <70% or Cohen's kappa < 0.2 were removed, resulting in a total of 77 items for the EAT-HC [16]. The majority of the instrument items are dichotomous and require a "Yes" or "No" response ($n = 27$). For 13 items, selection based on a three- to four-point categorical response scale is possible. In addition, for a number of items, the option "Not applicable" ($n = 32$) can be selected or an additional point awarded under Key Design Principle "Risk reduction" if the design of an environmental safety feature is unobtrusively applied ($n = 6$). For each subscale score, individual scores are added and converted into percentage results [14]. The EAT-HC was tested on a range of psychometric performance criteria. In a systematic review of tools for assessing the physical environment in health care, Elf et al. (2017) rated structural validity as fair and cross-cultural validity as good based on the COSMIN checklist [11,17]. With regard to content validity, the creators of the EAT-HC refer to the development of items based on a systematic review [8] and the EAT [9]. Pearson correlation was calculated between the overall score of the EAT-HC, the Therapeutic Environment Screening Survey for Nursing Homes (0.72), and the Special Care Unit Environmental Quality Scale (0.34) to evaluate concurrent validity ($p = 0.01$) [16]. Interrater reliability was calculated using the assessments of two unexperienced raters from a convenience sample of 30 nursing homes. The interclass correlation coefficients (ICCs) for each subscale varied between 0.52 and 0.99. Internal consistency was assessed with Cronbach's α values ranging from 0.57 to 0.88 and was interpreted as satisfactory [16]. The instrument was developed for application in research [18,19] and environmental consultation at Dementia Training Australia.

The EAT-HC was translated and culturally adapted for the German context as part of a project of the Deutsches Zentrum für Neurodegenerative Erkrankungen e.V. [13]. In accordance with the adapted versions of the EAT-HC for Singapore [20] and Japan [21], a multistep translation process of the instrument for the German context took place according to the guidelines of the World Health Organization (1998) [22]. Potential users of the future instrument from dementia-related health care research and dementia care were involved in the process. Additionally, an evaluation of content validity calculating the content validity index [23] twice took place as part of the adaptation phase. Content validity at the item level (I-CVI) varies between 0.33 and 1.0 regarding relevance in the German long-term care setting

and between 0.38 and 1.0 in terms of comprehensiveness. At the subscale level, the scores range between 0.69 and 0.99, and 0.57 and 0.97, respectively. Three items of Key Design Principle "Risk reduction" are controversial and will only be applied in secure living units in Germany [13]. Consequently, two versions of the instrument—for secure living units with 77 items and for nonsecure living units with 74 items—are available for the test procedure described in this article. Although the original Australian instrument shows satisfactory psychometric quality, the EAT-HC's performance needs to be rechecked after cross-cultural adaptation. Since no German gold standard for assessing the built environment exists, neither concurrent validity nor parallel test reliability can be determined. Consequently, this study focuses on interrater reliability. In addition, we assess acceptance of data collection, time-related aspects of the procedure and practicability. Polit and Beck (2017) summarize these aspects under the term "feasibility"—a smaller test before beginning with a larger research project [24]. This article follows the *Guidelines for Reporting Reliability and Agreement Studies* (GRRAS) [25].

2. Materials and Methods

2.1. Design

The study involved a feasibility and interrater reliability study. First, the feasibility study examined acceptance of data collection, the time-related aspect of the procedure and its practicability. Second, the reliability study involved testing interrater reliability at the item and subscale levels as well as internal consistency.

2.2. Sampling and Participants

The sample size for measuring interrater reliability was computed using the EAT-HC calculation according to the recommendations of Steiner and Norman (1995) and the methodological guidelines of Walter et al. (1998) [16,26,27]. As information on proportions used in the original Australian templates is not reported, models were calculated for different distributions of characteristics.

For the feasibility study, three nursing homes from the first author's professional network in one urban area and two rural areas participated. For the reliability study, we used data from a convenience sample of living units: 170 nursing homes within a 20 km radius of the research institute received a request to participate in the study. Upon initial contact, we checked whether the facilities had a secure living unit, as we focused on testing all 77 items of the G-EAT. If the requested facilities had such a living unit, this unit was included in the study. If all units appeared to be open, the facility managers decided for a given living unit how the assessment could least disturb the residents' daily routines.

2.3. Instruments

2.3.1. German Environmental Audit Tool

The German Environmental Audit Tool—High Care (G-EAT) is the adapted version of the above characterized EAT-HC [16]. Deviating from the original, we apply two versions of the G-EAT in this study: a version for nonsecure living units that contains 74 items and a 77-item version for secure living units [13].

2.3.2. G-EAT Context Questionnaire

This questionnaire contains different items regarding (a) structure, (b) architecture, (c) financing, (d) staff, (e) residents and (f) meals. Some items have been applied in the study DemenzMonitor [28], and others are newly developed and were tested within the feasibility study. We applied this questionnaire to describe the context of the participating living units, but it did not serve as part of the interrater reliability testing.

2.3.3. Additional Measurements to Test Feasibility

To assess the feasibility of the G-EAT, we measured the time spent explaining, assessing and reflecting the assessment with staff members in nursing homes. We focused on the

Int. J. Environ. Res. Public Health **2022**, *19*, 1050

4 of 15

well-being of the residents upon assessing and entering their living units. Additionally, practicability in terms of the tool's format was checked with a paper-pencil version versus a digital version on a tablet.

2.4. Data Collection

To test feasibility, the G-EAT was conducted by two researchers (AF and RP) and at least one staff member of the respective nursing home. Allied staff members were familiar with the living unit to assess the group of residents. First, the G-EAT and Key Design Principles were introduced, and then the staff member was asked to show the living unit, additional spaces used by the group of residents and the outdoor area. Afterward, all raters—from research and practice—created a common definition of the living unit to delineate the boundaries of the unit. Every rater filled in the G-EAT independently while walking around the living unit. The researchers took field notes when they identified difficulties in terms of disturbing residents or staff through data collection (e.g., during meals in common areas, group activities or conversations). After the assessment, questions on the G-EAT that could not be answered by the researchers themselves were discussed with the staff member (e.g., regarding residents' rooms). In addition, the staff member was asked about his or her experiences regarding disrupting the daily routine of the residents. Results from the feasibility study were included through team discussion processes into the interrater reliability data collection.

To test interrater reliability, data collection took place from August to December 2019 at the participating nursing homes. Two raters (AF and KS) with a background in nursing science as well as training as registered nurses conducted the assessment. Both were familiar with the instrument from various training sessions. In addition, the creators of the EAT-HC had trained AF during the instrument adaptation process. To allow independence of assessment, raters did not talk to each other about their assessments during data collection and, if possible, started their assessments in different areas in the living unit.

To avoid interrupting the residents' meals, assessments were conducted between 9:00 a.m. and 11:30 a.m., and 1:00 p.m. and 4:30 p.m. Moreover, one to three staff members of the facility completed the instrument to further evaluate its feasibility. The results of their perspectives are reported elsewhere. After obtaining informed consent and explaining the project data, the scientific raters illustrated data collection with the G-EAT. Subsequently, the staff members guided the researchers through the living units, and all raters—from research and practice—found a common definition of the rooms to be visited. Raters were told to assess the built environment with the G-EAT independently and not discuss the items during data collection. If items remained open or questions could not be answered due to ethical considerations, the answers from the staff members were used. For these items, we are not able to compute the ICCs but the proportions of agreement.

2.5. Statistical Analysis

Interrater reliability was computed at the item and subscale levels. At the item level, unweighted Cohen kappa was used to calculate rater agreement [29]. At the subscale level, intraclass correlation coefficients were applied. In accordance with the McGraw and Wong convention (1996) [30], we utilize a formula to calculate *"two-may mixed effects, absolute agreement and single rater/agreement"* (Koo and Mae, 2016, p. 157) [31]. For internal consistency, we computed Cronbach's alpha. In addition, proportions of agreement for all items were calculated (see Appendix A). Analyses were carried out using SPSS Version 25 (New York, NY, USA) [32].

2.6. Ethical Approval

The Ethics Commission of the German Society for Nursing Science approved this study (proposal number: 18-005). A letter from the study team informed staff and residents about data collection. Researchers only assessed shared spaces in the living unit and did not enter resident rooms without their invitation. Allying facility managers and staff members

Int. J. Environ. Res. Public Health **2022**, *19*, 1050

5 of 15

were asked to break up times of data collection if they sensed that the research staff would disrupt the daily routine of the residents.

3. Results

3.1. Participating Living Units

Forty nursing homes in the federal state of North Rhine-Westphalia participated in the interrater reliability study. The context questionnaire was only completed by a subset of the living units (*n* = 33). Table 1 shows their main characteristics: most were based in communities with 100,000 or more inhabitants (81.3%); only one facility was based in a more rural region (less than 20,000 inhabitants). The nursing homes varied in their numbers of units (2–10) and places (42–250). Only one unit of every nursing home was included. A total of 66.7% of the living units had an integrative living concept where residents with and without dementia lived together. Some living units were divided into smaller living groups by organization (37.5%) and/or spatially. In these cases, raters decided if they agreed with this from a spatial layout perspective, so only two living units were assessed with two living groups each.

Table 1. Structural characteristics of the included living units in stage 2.

Characteristics (*n* = 33)	Sample
	%/M (N/Range)
Sponsorship	
nonprofit	67 (22)
profit	33 (11)
Size of the community *	
<20,000 inhabitants	3 (1)
20,000–100,000 inhabitants	16 (5)
100,000–1,000,000 inhabitants	81 (26)
Structure of the facility	
Number of units	4 (2–10)
Number of places (full-time)	96 (42–250)
Year of construction *	
<1945	3 (1)
1945–1959	6 (2)
1960–1979	22 (7)
1980–1999	19 (6)
2000–2010	28 (9)
>2010	22 (7)
Time of last rebuilding	
No rebuilding	27 (9)
Over the last 2 years	15 (5)
3–5 years ago	9 (3)
6–10 years ago	15 (5)
More than 10 years ago	27 (9)
Unknown time period	6 (2)
Characteristics of included living units	
Integrative living concept	67 (22)
Division into living groups	36 (12)
Number of resident rooms	25 (10–45)

* missing value (*n* = 1).

Regarding architectural history, living units showed different characteristics: most had never been rebuilt (27.3%), or rebuilding had occurred more than 10 years ago (27.3%). The year of construction varied from 1933 to 2019.

3.2. Feasibility

The question regarding the practicability of the G-EAT assessment revealed that a paper-pencil version rather than a digital version was preferable for the IRR test. Comments and remarks on individual items can be added more easily, and sketches of living units or particular architectural features can be drawn on paper.

Acceptance of the data collection may be divided between residents' and staff's perceptions. Observations of the researchers and responses of the staff did not indicate that the presence of the researchers in the living unit assessing with the G-EAT disturbed the residents. With regard to their own acceptance, staff reported that they considered the questionnaire to be very long and that some items were difficult to interpret.

The duration of the assessments varied between 30 and 90 min (time from entering the living unit to all questions being answered). The length of the assessment not only depended on the size of the living unit but also on the related common areas (outdoor area, cafeteria, chapel, etc.) and their locations in the building. The use of these spaces at the time of assessment also mattered. For example, we did not enter commonly used rooms while residents ate lunch. Based on this knowledge, a period was chosen for IRRs during which residents did not take meals to avoid interrupting meals.

3.3. Interrater Reliability

For interrater reliability at the subscale level, ICC could be calculated for 8 out of 10 key design principles (see Table 2). The first key design principle cannot be calculated because of its qualitative nature. Key Design Principle II cannot be analyzed due to the small number of items involved ($n = 2$), so conclusions about reliability can be drawn only on an item level. ICCs range from 0.662 (III) to 0.869 (IV). According to Koo et al. (2016), four subscales showed moderate reliability, and two showed good reliability [31]. At the item level, Cohen's kappa could be calculated for 63 items (see Appendix A). Items of Key Design Principle *"Environmental design as part of the care philosophy"* were not included, as agreement is the intended outcome. In addition, nine items could not be calculated due to a statistical problem, as one or both raters always chose one answer possibility. According to Landis and Koch (1977), 6 items showed poor agreement (K < 0.00), 6 showed slight agreement (K 0.00–0.20), 12 showed fair agreement (0.20–0.40), 9 showed moderate agreement (K 0.40–0.60), 22 showed substantial agreement (0.60–0.80) and 8 showed almost perfect agreement (K > 0.8). [33].

Table 2. Interclass correlation coefficients on German Environment Audit Tool (G-EAT) subscale.

No.	Key Design Principle *	ICC (CI$_{95\%}$)	*p* Value	$_N$Items	Interpretation of ICC Value **
II	Provide a human scale	-	-	2	
III	Reduce risks unobtrusively	0.662 (0.452–0.803)	<0.001	17	Moderate reliability
IV	Allow people to see and be seen	0.869 (0.769–0.927)	<0.001	10	Good reliability
V + VI [1]	Manage levels of stimulation	0.728 (0.539–0.846)	<0.001	26	Moderate reliability
VII	Support movement and engagement	0.730 (0.505–0.854)	<0.001	9	Moderate reliability
VIII	Create a familiar place	0.698 (0.504–0.825)	<0.001	4	Moderate reliability
IX + X [2]	Links to the community	0.712 (0.516–0.835)	<0.001	9	Moderate reliability

* according to Fleming & Bennett (2015) [16]; ** according to Koo et al. (2016) [31]; [1] contains two KDPs: "Reduce negative stimulus" and "Enhance positive stimulus"; [2] contains two KDPs: "In the living unit" and "In the community".

3.4. Internal Consistency

Cronbach's alpha could not be computed for Key Design Principle I (qualitative character) and Key Design Principle II (small number of items). For the other subscales, scores varied from 0.362 (VII) (poor reliability) to 0.688 (IV) moderate reliability) [31] (see Table 3).

Table 3. Internal consistency of the G-EAT.

No.	KGP	$_N$Living Units	Cronbach's α	$_N$Items
II	Provide a human scale	41	-	2
III	Reduce risks unobtrusively	41	0.568	13
IV	Allow people to see and be seen	42	0.688	10
V + VI [1]	Manage levels of stimulation	41	0.353	25
VII	Support movement and engagement	42	0.362	9
VIII	Create a familiar place	42	0.503	4
IX + X [2]	Links to the community	42	0.521	9

[1] contains two KDPs: "Reduce negative stimulus" and "Enhance positive stimulus"; [2] contains two KDPs: "In the living unit" and "In the community".

3.5. Modifications According to Psychometric Properties

To modify the items, we used the interrater reliability values as well as previously collected data on content validity [13] and information about practicability from the field, which were obtained from field notes of the first and second authors. Nineteen items were revised, and the gender form was simplified across the items due to the German division between female, male and diverse notation. Modifications of the items are based on several team discussions; initially, the first author suggested alternative formulations according to the test results. These were discussed with the other authors in different group sessions until a consensus was achieved. For three items, we were not able to find reformulations that could address their poor psychometric properties, so they were removed. Additionally, some items showed poor reliability, as they covered more than one aspect. We divided these and further developed three new items addressing culturally specific needs in the field of dementia-specific care detected during data collection (see Table 4). For example, an item measuring the possibility for overnight stays from family members at nursing homes was raised several times and seems to be an important factor in helping connect families.

Table 4. Overview of main item modifications.

Key Design Principle	Removed Items
III	Are different corridors clearly recognizable so residents can identify where they are?
III	Is the bed placed or can it be placed so that from lying down, the toilet seat can be seen?
V	Does each room have a distinctive character and atmosphere?
	New items
III	Can the exit leading to the outdoor area be seen from the dining room?
III	Do lying residents have a view to the outside from the dining room?
VII	Is there a clearly defined path to the outdoor area that avoids dead ends and locked exits?
VII	Are there sunny areas along the path in the outdoor area?
VII	Is there a shaded seating area in the immediate surrounding of the facility?
IX	Is there a space for private conversations in the living unit?
X	Is there a room within the facility for families to stay overnight?

4. Discussion

Our study aimed to test the feasibility, interrater reliability and internal consistency of the German Environmental Audit Tool. We found that the instrument is feasible in a paper-pencil version and that additional information according to the boundaries of the living unit is necessary before assessment. At the subscale level, the G-EAT demonstrated moderate to good reliability, and items with poor kappa values were discussed and modified. Internal consistency varies between subscales but with overarching room for improvement in item categorization. The identified test-theoretical weaknesses in the reliability of the G-EAT require reflection on the content and the different cultural backgrounds of the original instrument and its adapted version.

Int. J. Environ. Res. Public Health **2022**, 19, 1050

8 of 15

5. Reflecting Cultural Differences in Germany and Australia

A number of poor reliability scores were derived from the questionable validity of the items in the German setting. Although the G-EAT received extensive cultural adaptation during translation—involving experts from the field and research [13]—the interrater reliability values and contextual data indicate that some questions require further semantic and linguistic adaptation. In addition, ceiling effects, especially for items of Key Design Principles *"Support movement and engagement"* and *"Reduce risks unobtrusively"*, suggest that the differentiation of questions or a Likert scale [34] with more precise answers would be useful for the German setting. For example, the question of whether acoustic stimuli are used outside could be answered with "yes" for almost all living units. However, variance between the sound of a forest and a large-scale bird fountain could not be mapped with this tool. These results are supported by research on other cross-cultural adaptations of the G-EAT. Brennan et al. (2021) report for the adapted Japanese version (EAT-HC-JV) that cultural design elements such as foot baths or a tea ceremony room cannot be depicted with the original list of items [21]. Similar findings are reported by Sun and Fleming (2021) for the Singaporean version. To address this challenge, the author included additional items for the areas of palliative care, technology and spirituality in the Singapore Environmental Assessment Tool [20].

6. Addressing the G-EAT Characteristics for Research Purposes

Other items seem to not be applicable in our study due to their composition. For example, some questions ask for two constructs (e.g., orientation and stimuli), while others contain two questions (e.g., view of outside from living and dining rooms). This contradicts the methods of questionnaire construction [35]. Before testing interrater reliability, we did not change any of these items, as we aimed to keep the German questionnaire as similar as possible to the original Australian version to ensure cross-cultural comparisons. However, as the G-EAT shall provide precise conclusions regarding the built environment for research and health care teams at nursing homes, there is a need for further improvement and thus also to move away from the original instrument. The need to do so is underpinned by the results of an instrument evaluation by Quirke et al. (2021). The authors compared the Environmental Audit Tool (EAT) (previous version of the EAT-HC) to two other instruments concerning their use for planning, detailing and managing the built environment. The authors found that the EAT primarily (60%) facilitates the planning of long-term care facilities and only indirectly (12%) post-occupancy—the aspect most applicable to the purposes we focus on [36]. We initiated the required development of the G-EAT by removing inapplicable items and reformulating other cultural-specific questions to help draw a more accurate and complete map of the built environment. We also anticipate increasing the internal consistency of the data. However, this should be verified by appropriate procedures such as exploratory factor or Rasch analysis in the future.

7. Enhancing Understanding of the Dementia-Specific Design

In our study, two raters who were familiar with the G-EAT construct and with the relevance of the dementia-specific environment assessed the built environment. Nonetheless, the data collection results show that additional information should be provided to enable a more prescient response and thus interpretation of the G-EAT in future applications in research and long-term care practice.

Apart from the new items already presented and those that had to be removed, we modified the G-EAT to facilitate the interpretation of the items. For instance, all items were checked with regard to their readability, and phrasing-nested sentences that are inherent to the German language were reformulated to improve the items' meaning. Furthermore, two additional elements were added to each item: notes on data collection and a comment field. Under *"Notes on data collection,"* we formulated instructions on how to answer an item on where the rater should stand when assessing visual axes. For example, acoustic stimuli in indoor areas are not applied, which might prevent stimuli from overwhelming

residents [37]. Background information was added to describe the meaning of the question in more detail. This necessity is also pointed out by Sun (2021) and Brennan and colleagues (2021), who emphasize the different literacy levels of staff of facilities using the instrument to collect data [38]. Even for other instruments that depict the built environment, additional information is provided in manuals to address the difficulties of this difficult-to-capture construct [39–41].

8. Built vs. Social Environment

In addition to the cross-cultural and construct-related challenges described above, the fluctuating reliability values may also reflect a conceptual problem. Some points of raters' disagreement arose due to raters attributing certain environmental aspects more to the architecture of the facility (built environment) or to the concept of the living unit (social environment). For example, a radio in a living unit can be mentioned here: it can offer a positive acoustic stimulus if it plays music known to the residents. To initiate this stimulus, staff usually turn on the radio, which represents a kind of dementia-specific intervention or serves as the beginning of a music group activity. The fact that these intersections between built and social environments have impacts and potential for people with dementia at home has already been explored [42]. Sun (2020) also reported that raters attribute or exclude different aspects of the built environment when testing the S-EAT [43]. This evidence from the literature shows that intersectionality—which represents an insurmountable challenge for test theory—cannot be neglected for dementia-specific care. Regarding G-EAT application, this means that a rater may allocate an unfulfilled item to the social environment and thus not systematically record it even though it is implemented.

9. Limitations of the Study

The study outlined here presents a few limitations that need to be considered when interpreting the results. We did not achieve the calculated sample size for all items, as staff and time resources did not permit the completion of such a comprehensive survey; hence, the sample size was set at 42 living units to ensure data collection for practical research reasons. As the results of this study must be considered against the clinical background, we may apply Cicchetti's (1999) methodological recommendation that *"one should simply [. . .] pay attention to the type of subject being assessed [and] train the examiners"* (p. 570) [44] if the intended sample size cannot be achieved. In addition, after data collection, the raters reflected on situations experienced in the field to assess the ethical aspects of assessing spaces while residents were present. Interference with the subsequent assessments could not be excluded but will be analyzed in more detail elsewhere. Since the general conditions for building and managing nursing homes vary between federal states, only nursing homes in one federal state of Germany participated. The different conditions in other states might affect the results of the G-EAT differently than those in the current sample. We also could not systematically capture a relevant source of bias in our study: seasonal changes in the environment. We conducted our surveys in different seasons but not repeatedly in the same living units. To measure the change sensitivity and interrater reliability of the G-EAT, a repeated-measures design—for example, as used by Calkins et al. (2007) [45]—would be useful in future studies.

Nonetheless, the study has strengths that should also be emphasized. Our comprehensive data collection and refinement of items based on additional data beyond interrater reliability values highlight that poor reliability often resulted from a lack of validity. This allows for this problem to be addressed without removing numerous items, thereby running the risk of altering or no longer being able to represent the underlying construct.

10. Conclusions

From the testing described in this paper, the German Environmental Audit Tool now exists in a modified version and may be applied in health care research. For our purpose of testing the G-EAT for its subsequent scientific use, our raters stood in for the population

of interest as individuals with prior knowledge of the instrument's content and handling. The usability and quality of the results for long-term care practitioners must be further evaluated with respect to the relevance of environmental design for practice and regarding the inclusion of this group in data collection. Given this background, we also plan to test the validity of the four new items in upcoming studies and intend to work on improving the internal consistency of the instrument with the implementation of the G-EAT in German health care research. Furthermore, the intersection of built and social environments poses a challenge to accurate data collection that should be considered in the course of the further application of the G-EAT in future surveys.

Int. J. Environ. Res. Public Health **2022**, *19*, 1050

11 of 15

Appendix A

Table 1. Cohen's Kappa and proportions of answers on G-EAT item-level.

KDP	Item	Cohens K (95% CI)	Strength of Agreement [+]	Proportions of Answers in %	
				NO	YES
II	Common areas are comfortable in scale	−0.077 (−0.148–(−0.006))	poor	9.8	90.2
III	Outside, access is step-free	0.352 (0.087–0.617)	fair	22.0	78.0
III	Outside, floor surfaces are safe	0.556 (0.313–0.799)	moderate	25.6	74.4
III	Outside, path surfaces are even	0.381 (0.149–0.613)	fair	46.3	53.7
III	Outside, paths are obstacle-free	−0.040 (−0.094–0.014)	poor	6.1	93.9
III	Inside, floor surfaces are safe	0.140 (−0.176–0.456)	slight	11.0	89.0
III	Inside, contrast between floor surfaces is avoided	0.481 (0.228–0.734)	moderate	23.2	76.8
III	Inside, ramps are wheelchair accessible	/ *		1.2	98.8
III	Bed/ensuite transfer is easy	/ *		1.2	98.8
IV	Lounge room is seen by staff	0.618 (0.342–0.894)	substantial	13.4	86.6
V	Doors to dangerous areas are seen	0.105 (−0.152–0.362)	slight	32.9	67.1
V	Wardrobes are cluttered	1	almost perfect	80.2	19.8
V	Public address/paging/call system is intrusive	0.829 (0.638–1)	almost perfect	17.1	82.9
V	Doors are noisy when closing	/ *		98.8	1.2
V	Visual clutter is absent	0.607 (0.37–0.844)	moderate	78.0	22.0
V	Inside, glare is avoided	0.481 (−0.022–0.984)	moderate	96.3	3.7
VI	Rooms are easily identifiable	−0.068 (−0.124–(−0.012))	poor	92.7	7.3
VI	Dining room is clearly recognizable	−0.043 (−0.107–0.021)	poor	90.2	9.8
VI	Toilet pan can be seen from bed	0.548 (0.222–0.874)	moderate	87.8	12.2
VI	Inside, contrast aids visibility of surfaces/objects	/ *		6.1	93.9
VI	Inside, olfactory cues are used	0.381 (0.159–0.603)	fair	57.3	42.7
VI	Inside, tactile cues are used	−0.040 (−0.094–0.014)	poor	6.1	93.9
VI	Inside, auditory cues are used	0.313 (0.132–0.494)	fair	51.2	48.8
VI	Outside, contrast aids visibility of surfaces/objects	/ *		3.7	96.3
VI	Outside, materials/finishes are varied	0.656 (0.131–1)	substantial	3.7	96.3
VI	Outside, olfactory cues are used	1	almost perfect	2.4	97.6
VI	Outside, auditory cues are used	0.222 (−0.036–0.48)	fair	31.7	68.3
VI	Outside view from dining/lounge is attractive	0.045 (−0.212–0.302)	slight	14.6	85.4

Table 1. *Cont.*

KDP	Item	Cohens K (95% CI)	Strength of Agreement [+]	Proportions of Answers in %		
VII	In-/outside path clearly returns residents to starting point	0.049 (−0.211–0.309)	slight		79.3	20.7
VII	Outside, path passes participation opportunities	0.634 (0.392–0.876)	substantial		20.7	79.3
VII	Outside, activity choices are available	0.243 (0.044–0.442)	fair		22.0	78.0
VII	Outside, seating is available	0.774 (0.524–1.024)	substantial		12.2	87.8
VII	Outside, sunny and shady areas are available	0.548 (0.222–0.874)	moderate		12.2	87.8
VII	Outside, passive activities are available	/ *			1.2	98.8
VII	Outside, verandas and shaded seating are available	/ *			0.0	100.0
VII	Inside, path passes participation opportunities	0.189 (−0.096–0.474)	slight		19.5	80.5
VII	Inside, path passes conversation/rest areas	0.598 (0.335–0.861)	moderate		18.3	81.7
IX	Dining room allows for dining alone	0.757 (0.538–0.976)	substantial		18.3	81.7
IX	Lounge room includes private conversation areas	0.217 (−0.007–0.441)	fair		20.7	79.3
IX	Outside, private conversation areas are available	/ *			1.2	98.8
X	Community interaction areas are accessible	1	almost perfect		2.4	97.6
X	Family/dining area is available in facility	/ *			0.0	100.0
X	Visitor break area is available	0.643 (0.265–1)	substantial		6.3	93.8
				N/A	NO	YES
III	Outside, paths have appropriate width	0.652 (0.458–0.846)	substantial	0.0	59.8	40.2
III	Outside, ramps are wheelchair accessible	0.494 (0.209–0.779)	moderate	0.0	17.1	82.9
IV	Garden/outside area exit is seen from lounge/dining room	0.847 (0.707–0.987)	almost perfect	0.0	63.4	36.6
IV	Dining room is seen from lounge room	0.919 (0.787–1)	almost perfect	0.0	18.3	81.7
IV	Toilet is seen from lounge room	0.692 (0.504–0.88)	substantial	0.0	63.4	36.6
IV	Toilet is seen from dining room	0.754 (0.565–0.943)	substantial	0.0	73.2	26.8
IV	Dining room is seen by staff	0.639 (0.407–0.871)	substantial	0.0	19.5	80.5
IV	Outside, resident area is seen by staff	0.377 (0.153–0.601)	fair	0.0	73.2	26.8
VI	Lounge room is clearly recognizable	−0.024 (−0.258–0.21)	poor	1.2	23.2	75.6
VI	Corridors are clearly identifiable	0.196 (−0.018–0.41)	slight	7.3	28.0	64.6

Table 1. *Cont.*

KDP	Item	Cohens K (95% CI)	Strength of Agreement [+]	Proportions of Answers in %			
VI	Bedrooms are individually identified	0.721 (0.512–0.93)	substantial	0.0	20.7	79.3	
VI	Shared bathrooms/toilets are clearly identified	0.731 (0.576–0.886)	substantial	31.7	18.3	50.0	
				N/A	NO	YES	UNOB. **
III	Resident kitchen has safe appliances	0.641 (0.491–0.791)	substantial	16.0	28.4	30.9	24.7
III	Resident kitchen has master switch	0.789 (0.664–0.914)	substantial	17.1	43.9	9.8	29.3
				0–25%	26–50%	51–75%	76–100%
IV	Lounge room is seen from bedrooms	0.740 (0.6–0.88)	substantial	52.4	13.4	12.2	22.0
				N/A	MANY	A FEW	NONE
VIII	Lounge furniture is familiar	0.386 (0.124–0.648)	fair	0.0	0.0	26.8	73.2
					MANY	A FEW	NONE
VIII	Bedroom furniture is familiar	0.711 (0.514–0.908)	substantial		2.4	24.4	73.2
VIII	Bedrooms have residents' own decorations/photos	1	almost perfect		95.1	4.9	0.0
VIII	Bedrooms have residents' own furniture	0.786 (0.618–0.954)	substantial		65.9	34.1	0.0
				NO	1	2 OR MORE	
IX	Inside, small group areas are available	0.376 (−0.07–0.822)	fair	0.0	4.9	95.1	
				NO	1	2	3 OR MORE
IX	Inside, private conversation areas are available	0.658 (0.487–0.829)	substantial	0.0	9.8	30.5	59.8
					1	2 OR 3	4 OR MORE
IX	Inside, variety of different areas are available	0.218 (−0.006–0.442)	fair		12.2	57.3	30.5
				0–25%	26–50%	51–75%	76–100%
VI	Pathway is defined from bedroom to dining room	0.441 (0.268–0.614)	moderate	52.4	19.5	12.2	15.9
IV	Bedrooms are seen from lounge room	0.497 (0.303–0.691)	moderate	62.2	30.5	3.7	3.7
IV	Dining room is seen from bedrooms	0.746 (0.601–0.891)	substantial	62.2	13.4	9.8	14.6
VI	Window view is attractive from bed	0.398 (0.217–0.579)	fair	1.2	9.8	24.4	64.6
				0–25%	26–74%	75–100%	
VI	Toilet seats contrast with background	0.759 (0.598–0.92)	substantial	58.5	0.0	41.5	
				30+	17–29	11–16	≥10
II	Number of residents in the unit	1	almost perfect	12.2	45.1	40.2	2.4

* Cohen's Kappa not computable; ** yes, unobtrusively. [+] according to Landis & Koch (1977) [33].

Int. J. Environ. Res. Public Health **2022**, *19*, 1050

14 of 15

References

1. Statistisches Bundesamt. Pflege. Pflegebedürftige nach Versorgungsart, Geschlecht und Pflegegrade. 2019. Available online: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Pflege/Tabellen/pflegebeduerftige-pflegestufe.html;jsessionid=0A7917CDD4AF7C7945B9C27B8977F9CB.live732 (accessed on 30 September 2021).
2. Gesundheitsberichterstattung des Bundes. Pflegeheime und Verfügbare Plätze in Pflegeheimen. Gliederungsmerkmale: Jahre, Region, Art der Einrichtungen/Plätze, Träger. 2019. Available online: https://www.gbe-bund.de/gbe/pkg_isgbe5.prc_menu_olap?p_uid=gastd&p_aid=45814573&p_sprache=D&p_help=2&p_indnr=570&p_indsp=&p_ityp=H&p_fid= (accessed on 30 September 2021).
3. Hoffmann, F.; Kaduszkiewicz, H.; Glaeske, G.; Bussche, H.V.D.; Koller, D. Prevalence of dementia in nursing home and community-dwelling older adults in Germany. *Aging Clin. Exp. Res.* **2014**, *26*, 555–559. [CrossRef] [PubMed]
4. Schaufele, M.; Köhler, L.; Hendlmeier, I.; Hoell, A.; Hoell, A. Prevalence of dementia and medical care in German nursing homes: A nationally representative survey. *Psychiatr. Prax.* **2013**, *40*, 200–206.
5. Chaudhury, H.; Cooke, H.A.; Cowie, M.H.; Razaghi, B.L. The Influence of the Physical Environment on Residents With Dementia in Long-Term Care Settings: A Review of the Empirical Literature. *Gerontologist* **2018**, *58*, e325–e337. [CrossRef] [PubMed]
6. Woodbridge, R.; Sullivan, M.P.; Harding, E.; Crutch, S.; Gilhooly, K.J.; Gilhooly, M.L.M.; McIntyre, A.; Wilson, L. Use of the physical environment to support everday activities for people with dementia: A systematic review. *Dementia* **2018**, *17*, 533–572. [CrossRef]
7. Harrison, S.L.; Dyer, S.M.; Laver, K.; Milte, R.; Fleming, R.; Crotty, M. Physical environmental designs in residential care to improve quality of life of older people. *Cochrane Database Syst. Rev.* **2017**, *2017*, 1–17. [CrossRef]
8. Fleming, R.; Purandare, N. Long-term care for people with dementia: Environmental design guidelines. *Int. Psychogeriatr.* **2010**, *22*, 1084–1096. [CrossRef]
9. Fleming, R. An environmental audit tool suitable for use in homelike facilities for people with dementia. *Australas. J. Aging* **2011**, *30*, 108–112. [CrossRef]
10. Chaudhury, H.; Cooke, H. Design matters in dementia care: The role of the physical environment in dementia care settings. In *Excellence in Dementia Care*; Downs, M., Bowers, B., Eds.; Open University Press: London, UK, 2014; pp. 144–158.
11. Elf, M.; Nordin, S.; Wijk, H.; McKee, K.J. A systematic review of the psychometric properties of instruments for assessing the quality of the physical environment in healthcare. *J. Adv. Nurs.* **2017**, *73*, 2796–2816. [CrossRef]
12. Fleming, R.; Bennett, K. Environmental Audit Tool—Higher Care (EAT-HC) Handbook. In *Resource 4—Environmental Design Resources, Dementia Training Australia, Designing for People with Dementia*; University of Wollongong: New South Wales, Australia, 2017.
13. Fahsold, A.; Fleming, R.; Verbeek, H.; Holle, B.; Palm, R. German Translation, Linguistic Validation, and Cultural Adaptation of the Environmental Audit Tool—High Care. *Health Environ. Res. Des. J.* **2021**, 1–15. [CrossRef]
14. Smith, R.; Fleming, R.; Chenoweth, L.; Jeon, Y.-H.; Stein-Parbury, J.; Brodaty, H. Validation of the Environmental Audit Tool in both purpose-built and non-purpose-built dementia care settings. *Australas. J. Aging* **2012**, *31*, 159–163. [CrossRef]
15. Fleming, R.; Kelly, F.; Stillfried, G. 'I want to feel at home': Establishing what aspects of environmental design are important to people with dementia nearing the end of life. *BMC Palliat. Care* **2015**, *14*, 26. [CrossRef]
16. Fleming, R.; Bennett, K. Assessing the quality of environmental design in nursing homes for people with dementia: Development of a new tool. *Australas. J. Aging* **2015**, *34*, 191–194. [CrossRef] [PubMed]
17. Mokkink, L.B.; Terwee, C.B.; Patrick, D.L.; Alonso, J.; Stratford, P.; Knol, D.L.; Bouter, L.M.; De Vet, H.C.W. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual. Life Res.* **2010**, *19*, 539–549. [CrossRef] [PubMed]
18. Chenoweth, L.; Jeon, Y.-H.; Stein-Parbury, J.; Forbes, I.; Fleming, R.; Cook, J.; Cheah, S.; Fletcher, S.; Tinslay, L. PerCEN trial participant perspectives on the implementation and outcomes of person-centered dementia and environments. *Int. Psychogeriatr.* **2015**, *27*, 2045–2057. [CrossRef] [PubMed]
19. Low, L.-F.; Goodenough, B.; Fletcher, J.; Xu, K.; Casey, A.-N.; Chenoweth, L.; Fleming, R.; Spitzer, P.; Bell, J.-P.; Brodaty, H. The Effects of Humor Therapy on Nursing Home Residents Measured Using Observational Methods: The SMILE Cluster Randomized Trial. *J. Am. Med. Dir. Assoc.* **2014**, *15*, 564–569. [CrossRef]
20. Sun, J.; Fleming, R. The Development and Reliability of the Singaporean Environmental Assessment Tool (SEAT) for Facilities Providing High Levels of Care for People Living With Dementia. *Herd* **2021**, *14*, 289–300. [CrossRef] [PubMed]
21. Brennan, S.; Doan, T.; Bennett, K.; Hashimoto, Y.; Fleming, R. Japanese Translation and Validation of the Environmental Assessment Tool-Higher Care. *Herd* **2021**, *14*, 19375867211007856. [CrossRef] [PubMed]
22. World Health Organization. Programme on Mental Health, in WHOQOL User Manual. Available online: https://apps.who.int/iris/bitstream/handle/10665/77932/WHO_HIS_HSI_Rev.2?sequence=1 (accessed on 30 September 2021).
23. Lynn, M.R. Determination and Quantification Of Content Validity. *Nurs. Res.* **1986**, *35*, 382–385. [CrossRef]
24. Polit, D.F.; Beck, C.T. *Nursing Research. Generating and Assessing Evidence for Nursing Practice*, 10th ed.; Wolters Kluwer: Alfing, The Netherlands, 2017.
25. Kottner, J.; Gajewski, B.J.; Streiner, D.L. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *Int. J. Nurs. Stud.* **2011**, *48*, 659–660. [CrossRef]

Int. J. Environ. Res. Public Health **2022**, *19*, 1050

15 of 15

26. Streiner, D.L.; Norman, G.R. *Health Measurement Scales: A Practical Guide to their Development and Use*; Oxford University Press: New York, NY, USA, 1995.

27. Walter, S.D.; Eliasziw, M.; Donner, A. Sample size and optimal designs for reliability studies. *Stat. Med.* **1998**, *17*, 101–110. [CrossRef]

28. Palm, R.; Holle, B. Forschungsbericht der Studie DemenzMonitor. In *Umsetzung Demenzspezifischer Wohn—und Betreuungskonzepte in Einrichtungen der Stationären Altenhilfe*; Deutsches Zentrum Zentrum für Neurodegenerative Erkrankungen (DZNE) e.V.: Witten, Germany, 2016.

29. Mayer, H.; Nonn, C.; Osterbrink, J.; Evers, G. Qualitätskriterien von Assessmentinstrumenten—Cohen's Kappa als Maß der Interrater-Reliabilität (Teil 1)*. *Pflege* **2004**, *17*, 36–46. [CrossRef] [PubMed]

30. McGraw, K.O.; Wong, S.P. Forming Interferences About Some Intraclass Correlation Coefficients. *Psychol. Methods* **1996**, *1*, 30–46. [CrossRef]

31. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]

32. Corp, I. *IBM SPSS Statistics for Windows, Version 25.0*; IBM Group: New York, NY, USA, 2017.

33. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159. [CrossRef]

34. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.* **1932**, *22*, 55.

35. Boateng, G.O.; Neilands, T.B.; Frongillo, E.A.; Melgar-Quiñonez, H.R.; Young, S.L. Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front. Public Health* **2018**, *6*, 149. [CrossRef]

36. Quirke, M.; Ostwald, M.J.; Fleming, R.; Taylor, M.; Williams, A. Design stage evaluation tools for residential dementia care environments. *Facilities* **2021**, *39*, 828–842. [CrossRef]

37. Cohen-Mansfield, J.; Marx, S.; Thein, K.; Dakheel-Ali, M. The impact of stimuli on affect in persons with dementia. *J. Clin. Psychiatry* **2011**, *72*, 480–486.

38. Sun, J. The Appropriateness of Adapting the Australian Environmental Assessment Tool-High Care (EAT-HC) for Persons with Dementia in Singapore. *Herd* **2020**, *13*, 57–67. [CrossRef]

39. Agency for Integrated Care. *Looking to the Future: Designing and Managing Residential Facilities for People with Dementia*; Agency for Integrated Care: Singapore, 2021.

40. de Boer, B.; Beerens, H.C.; Katterbach, M.A.; Viduka, M.; Willemse, B.M.; Verbeek, H. The Physical Environment of Nursing Homes for People with Dementia: Traditional Nursing Homes, Small-Scale Living Facilities, and Green Care Farms. *Healthcare* **2018**, *6*, 137. [CrossRef]

41. Pholeros, P.; Bennett, K.; Welke, A.; Arch, M. Indigenous Aged Care Design Guide. Dement. Available online: https://www.healthabitat.com/wp-content/uploads/2020/04/DTA_R6_IACDG_DIGITAL.pdf (accessed on 21 August 2021).

42. Sturge, J.; Nordin, S.; Patil, D.S.; Jones, A.; Légaré, F.; Elf, M.; Meijering, L. Features of the social and built environment that contribute to the well-being of people with dementia who live at home: A scoping review. *Health Place* **2021**, *67*, 102483. [CrossRef] [PubMed]

43. Sun, J. Defining and assessing the characteristics of the built environment that contribute to the well-being of people with dementia living in aged care facilities in Singapore. In *School of Nursing*; University of Wollongong: New South Wales, Australia, 2020.

44. Cicchetti, D.V. Sample size requirements for increasing the precision of reliability estimates: Problems and proposed solutions. *J. Clin. Exp. Neuropsychol.* **1999**, *21*, 567–570. [CrossRef] [PubMed]

45. Calkins, M.; Szmerekovsky, J.G.; Biddle, S. Effect of Increased Time Spent Outdoors on Individuals with Dementia Residing in Nursing Homes. *J. Hous. Elder.* **2007**, *21*, 211–228. [CrossRef]