The Quality of Designs by Shape Grammar Systems and Architects:

A Comparative Test on Refurbishing Lisbon's *Rabo-de-Bacalhau* Apartments

Sara Eloy[1], Pieter Vermaas[2] and Marina Andrade[3]

[1] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisbon, Portugal
[2] Delft University of Technology, Delft, The Netherlands
[3] Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL, Lisbon, Portugal

Abstract:

This paper compares the quality of designs generated by shape grammar systems with designs created by professional architects. It describes an experiment in which evaluators rank the quality of preliminary designs for refurbishing Lisbon's *Rabo-de-Bacalhau* apartments, and that makes plausible that the design quality of the grammar solutions for this task is similar to that of the designs of the professional architects. It is argued that shape grammars can therefore be used as stand-alone design systems in architecture. The practical value is that shape grammar systems become alternative means to take up design challenges (such as refurbishing all of Lisbon's Rabo-de-Bacalhau apartments) that currently require too much effort of architects to be economically feasible.

1 Introduction

The promise of shape grammar systems is enriching architecture with algorithmic design processes that manipulate shapes rather than text or symbols, and that result in semantically and syntactically correct designs. This promise is generally taken as one being about the *tools* that support architects in their designing; the stronger understanding that shape grammars can be *stand-alone systems* that create designs independently of architects, is meeting scepticism about the quality of grammar solutions and raises concerns about the competition it introduces to architects (Flemming, 1994). The same scepticism exists about using algorithm systems and grammar systems for creating music, e.g., McDermott (2008), and art, e.g., March (1981), and for mimicking the creative processes involved in these fields, e.g., Jacob (1996). This paper is about the promise of shape grammar systems as stand-alone design systems. It identifies an architectural task that can be addressed by shape grammars and architects, and describes an experiment for comparing the design quality of the grammar solutions and architectural designs. The results of this experiment make plausible that the design quality of the grammar solutions can be similar to that of designs created by professional architects.

The architectural task we consider concerns the refurbishment[1] of apartments of the type "*Rabo-de-Bacalhau*" in the city of Lisbon. For this task a shape grammar is available (Eloy, 2012), hence can be used for comparing the design quality of grammar solutions and architect designs.

[1] The term 'refurbishment' is here used in the meaning of adjusting or rehabilitating apartments and buildings for improving their habitability socially, physically and functionally.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

A comparison by design quality suggests understanding what is meant with the quality of architectural designs for housing. This clarity could be achieved by giving a definition, drawing from the extensive literature on the topic (e.g., Ablem, *et al.*, 1979; Cabrita, *et al.*, 1998; Qualitel, 1998; Pedro, 2000; Natividad-Jesus, *et al.*, 2007; Maginn, *et al.*, 2008). In the experiment design quality is not defined but approached in an operationalistic way that allows quantitative analysis. Staying close to how it often is determined in architectural schools and competitions, it is assumed that design quality is a property of designs that can be determined through evaluation committees of qualified architects.

The results of the experiment make plausible the claim that the design quality of the grammar solutions for refurbishing Rabo-de-Bacalhau apartments is similar to that of designs created by professional architects. They therefore support the further position that there exist architectural tasks for which shape grammars can generate designs independently of architects yet with a quality similar to that of architects. The practical value of this result is that for some of these tasks shape grammar systems can be used as stand-alone design systems. This value lies not in replacing architects. For instance, for addressing more singular and specific tasks, commissioning an architect leads swifter and with less costs to a good design than by first developing a shape grammar and then applying it. The promise is rather for the opposite case of large-scale and standardised architectural tasks, such as refurbishing Lisbon's stock of Rabo-de-Bacalhau apartments. Such tasks when carried out by architects can soon become economically unfeasible for clients and be highly repetitive for the architects, whereas shape grammars are enabling architecture to take them up.

The plan for the paper is as follows. The next section introduces Rabo-de-Bacalhau apartments. Section 3 describes shape grammar systems and the grammar for refurbishing the Rabo-de-Bacalhau apartments. In Section 4 it is argued how architects and this grammar system can address the refurbishment task. Section 5 describes the experiment in which both professional architects and the grammar produced designs for this task. Section 6 gives the outcomes of the experiment consisting of evaluations of the quality of the resulting designs by an architectural evaluation committee. We analyse these outcomes in three steps. In Section 7 we argue that the aggregated quality of the designs generated by the grammar is similar to or even higher than the quality of the architect designs. In Section 8 we give a statistical proof that this first result holds also when the quality of the designs in the experiment was subject to random fluctuations. In Section 9 we show statistically that the design quality of the grammar solutions is similar to the quality of the designs as generated by two of the six professional architects participating in the experiment.

2 Lisbon's Rabo-de-Bacalhau apartments and their renovation

In the city of Lisbon, Portugal, there is a large stock of apartments referred to as Rabo-de-Bacalhau housing. These apartments were built in the 1940s and 1950s and are today candidate for renovation. The Portuguese term "*rabo-de-bacalhau*" means cod-tail in English and characterises the apartments through the similarity between the shape of a dried cod-fish and the apartments' projection with a rear wing narrower than the front one (Figure 1). Buildings containing Rabo-de-Bacalhau apartments have right and left displacements and their height varies between four to nine floors, with six the average. Their construction consists predominantly of a reticular concrete structure filled with masonry walls and with interior walls, usually with loadbearing characteristics.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

When focussing on typology and other features, Rabo-de-Bacalhau buildings come in four types (Figure 1). Yet these buildings display actually a wide variety of architectural differences due to differences in the dimensions of the urban plot in which they were built and in specificities of the interior layout. This means that Rabo-de-Bacalhau apartments may share a general overall floor-plan typology but do not come in a limited set of standardised configurations. A rough estimate is that there exist more than 30,000 Rabo-de-Bacalhau apartments.[2]

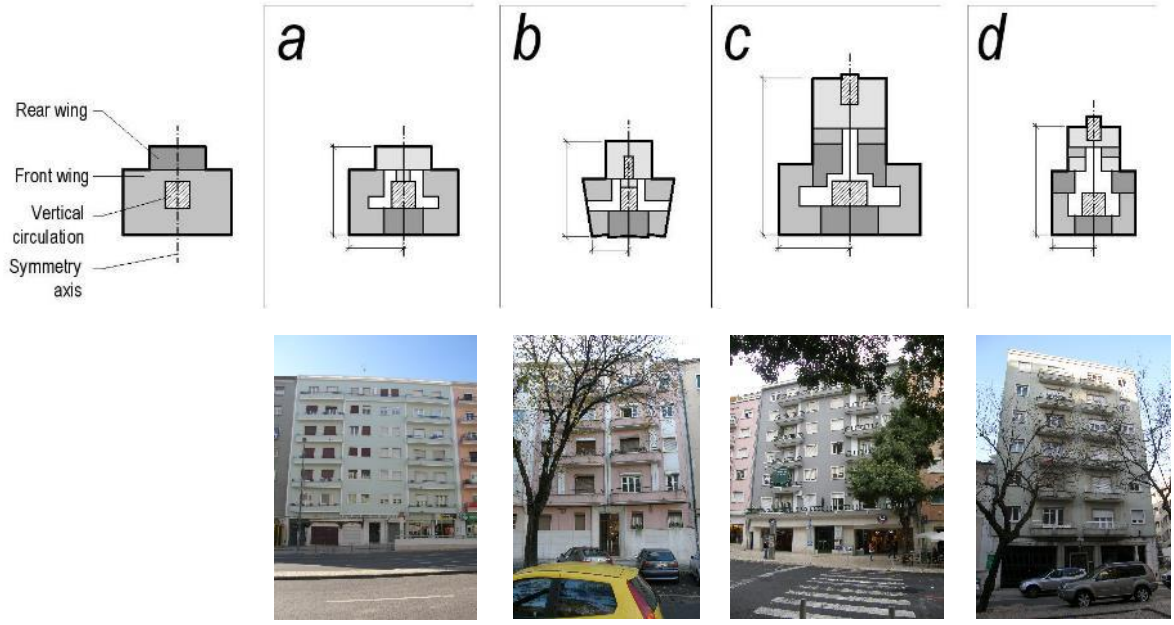

Figure 1: General floor plan and four Rabo-de-Bacalhau building types with example façade.

Rabo-de-Bacalhau buildings were originally owned by private developers who saw them as sources of income and rented the apartments to their inhabitants. Nowadays the apartments are essentially private property, individually or collectively, and used for residence or long-term rental.

Rabo-de-Bacalhau apartments are part of the continuity of the current social fabric of Lisbon and architectural characteristic to many of its neighbourhoods (Monteys, 2013). Yet their interior layouts reflect the social standards of Portuguese society of the 1950s, for instance, by containing rooms for in-house maids. Nowadays, building construction improvements are needed in Rabo-de-Bacalhau apartments, by their age, by a lack of maintenance and by incorrect interventions leading to structural fragilities. Moreover, functional adjustments are in order to update the interior layout to contemporary lifestyles.

There are two ways of addressing this renovation need: demolishing the Rabo-de-Bacalhau &buildings or refurbishing them. Demolishing and replacing the buildings is not preferable, since it would mean a major disruption of the mentioned social fabric of Lisbon by large-scale

[2] In Lisbon 12,000 buildings were constructed between 1946 and 1960 (INE, 2012), which falls within the 1940s-1950s period in which Rabo-de-Bacalhau buildings were built. Assuming that 1/5 of those 12,000 were Rabo-de-Bacalhau buildings, and that one building has as the average of 6 floors with two apartments (Eloy, 2012, p.134), then 28,800 apartments were built between 1946 and 1960. Since Rabo-de-Bacalhau buildings were also built before 1946 the true number of apartments is larger on this estimate.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

relocation of the current inhabitants, and since this action would contribute to the global phenomenon of spatial and morphological homogenisation of housing losing cultural specificities of cities (Lara and Youngchul, 2010). Demolishing is also not attractive from the point of view of ecological sustainability (Hall and Warm, 1994; Lees, 2008; Semes, 2009; Arthurson, 2012). Refurbishment does not have these disadvantages and seems therefore the better option; the city of Lisbon has actually explicitly endorsed refurbishment toward more sustainable urban management (CML, 2007).

In the next section we introduce a shape grammar for refurbishing Rabo-de-Bacalhau apartments, and in Section 4 we argue that this refurbishment has to be carried out per individual apartment for avoiding that also refurbishment leads to large-scale reallocation of inhabitants.

3 The Rabo-de-Bacalhau transformation grammar

Shape grammars are computational systems that allow the generation of design solutions based on shape rules. They were introduced in the 1970s by Stiny and Gips (1971), used initially in painting and sculpture and then in other areas such as architecture. Shape grammars come as analytical and original grammars (Stiny, 1976). *Analytical grammars* are used for analysing existing design languages and make the underlying design principles understandable, as for, for instance, Palladio villas (Stiny and Michell, 1978) and Hayat houses (Colakoglu, 2005). Analytical grammars enable the generation of existing designs within a design language, but also of new designs within the language. *Original grammars* are developed independently of existing design languages, and can be used for supporting new architectural designs by enabling architects to explore design alternatives instead of designing just one solution, and thus to create customised solutions. Research on original grammars has been focussed on customised urban settlements, housing and building construction processes (Sass, 2005; Tuncer and Sener, 2010; Duarte and Beirão, 2011).

Shape grammars can also transform existing designs, e.g., for generating design solutions for refurbishing housing. In (Eloy, 2012) such a grammar is given for Rabo-de-Bacalhau apartments. This grammar, called the *Rabo-de-Bacalhau transformation grammar* (RB-grammar, for short) is a computational generative design system that uses shape rules for adapting the layout of apartments to let them meet contemporary inhabitant wishes.

The Rabo-de-Bacalhau transformation grammar was developed on the basis of several refurbishing designs made by architects. These designs revealed patterns in the reasoning of the architects and the sequence of architectural actions by which they created refurbishment solutions. These patterns were captured by transformation rules for architectural elements in the Rabo-de-Bacalhau apartments, such as rules for moving walls of for changing the function of rooms. These rules finally make up the shape grammar (Eloy and Duarte, 2014). As any shape grammar the RB-grammar offers an increase in variety and customisation by generate multiple viable designs – for each specific client request and each specific original layout it delivers a different tailor-made solution (Figure 2) – without a corresponding increase in costs (Reffat, 2008). The RB-grammar uses a complex set of data that is cross-referenced in rules that encode knowledge about the aspects that are related to this type of architecture assignment, say, building information, required rooms and their characteristics, required constructional aspects and inhabitant's characteristics. By these rules the RB-grammar

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

transforms the spatial composition of an apartment to meet specified functional and constructional requirements. The grammar addresses the functional requirements by first identifying habitable characteristics and then intervening in the position of walls and in the connections or adjacencies between rooms. To address the constructional requirements the RB-grammar includes conditions on the demolishment and construction of walls and on the relocation of bathrooms and kitchen, such that the structural integrity of the building is maintained. The refurbishment methodology encompasses several steps that are applied in a hierarchical order. After defining the functional programme for a specific client, the RB-grammar starts by assigning housing functions room by room, in a predefined order, and changing shape (rooms and walls) whenever needed.

From an architectural point of view the advantage of the RB-grammar is that it generates solutions that respect the Rabo-de-Bacalhau building morphology and prevents structural fragilities. The fact that the system follows a single design language has in turn the disadvantage that it prevents alternative designs.

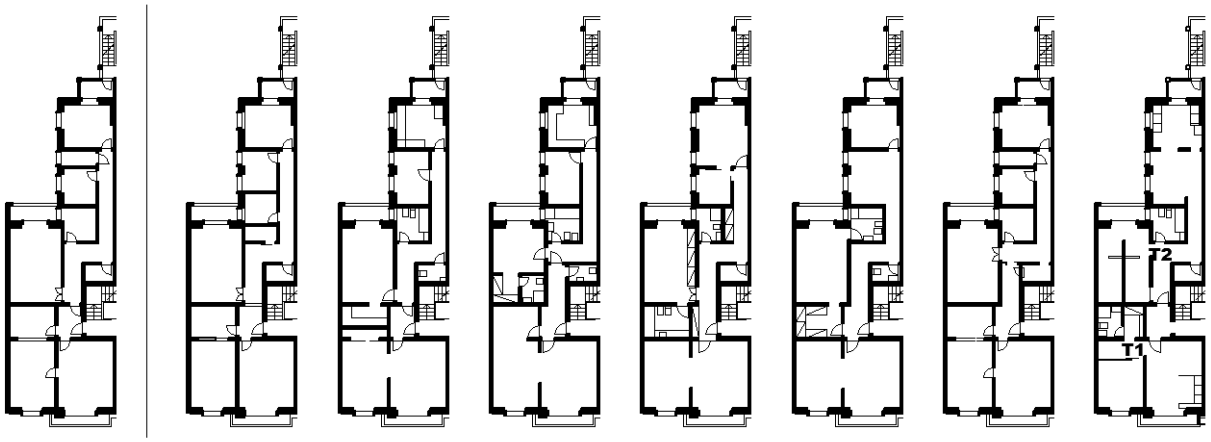

Figure 2: An original layout of a Rabo-de-Bacalhau apartment (far left) and seven different RB-grammar generated layouts using seven different client wishes.

4 Refurbishing Rabo-de-Bacalhau apartments with the RB-grammar

Architects can take up the task of refurbishing Rabo-de-Bacalhau apartments, and they can do so using the described Rabo-de-Bacalhau transformation grammar. Yet, this shape grammar could also be taken as a stand-alone system and generate refurbishment designs independently of architects. For assessing this possibility one could compare the solutions the RB-grammar generates with the designs (human) architects produce. It can be argued that a more realistic comparison is between the solutions the RB-grammar generates and preliminary designs that architects produce. In this argument it is assumed that the refurbishment concerns Lisbon's full stock of Rabo-de-Bacalhau apartments, should not become unrealistically expensive and should not reintroduce the disadvantage of large-scale reallocation of inhabitants.

Consider commissioning architects for refurbishing all Rabo-de-Bacalhau apartments. One option is that the architects develop a master plan consisting of a limited set of detailed refurbishment designs, where each design covers different requirements and wishes of the apartments' current inhabitants. Such a master plan provides flexibility for the inhabitants to

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

choose a design that fits their needs and wishes, allowing them to stay living in their apartments after refurbishment. This first option is however not possible due to the variety in Rabo-de-Bacalhau apartments (see Section 2); floor layouts are too diverse, meaning that the master plan should provide detailed refurbishment designs for each of the different layouts, multiplying the architectural workload and costs.

A second option is that architects give detailed refurbishment designs for clusters of apartments – say per Rabo-de-Bacalhau building or per a number of buildings of the same type located in a neighbourhood. In this case the design work is still multiplied by the number of clusters, which the architects may counter by creating per cluster only two or three detailed designs for the apartments. In this option there is not much room for the needs of individual inhabitants; they rather get to choose between the designs the architects offer, and if these designs do not meet the inhabitants' situation, they may decide not to return to their apartment after refurbishment. A cluster-wise refurbishment that does avoid this unwanted reallocation of inhabitants, loses the efficiency aimed at. In that alternative, defining a third option, the individual needs and wishes of inhabitants are taken into account, implying that architects give detailed refurbishment designs for each apartment individually, and making the enterprise truly labour intensive and expensive.

Hence, given the variety of floor plans of Rabo-de-Bacalhau apartments and given that refurbishment should not lead to substantial reallocations, architects have to find alternatives for limiting their workload when refurbishing apartments individually. One such alternative is the option that architects do not give detailed refurbishing designs but preliminary designs for each apartment on the basis of the inhabitants' wishes, and leave the detailing to other stakeholders, such as construction companies.

This fourth option can indeed be carried out by architects but also by the RB-grammar used as a stand-alone design system. And on this option the assessment of the RB-grammar as a stand-alone system consists of comparing the quality of the RB-grammar solutions with the quality of the preliminary designs that professional architects produce. The experiment we now describe focusses on this comparison.

5 The experiment

The experiment has the following structure. First, 3 assignments for refurbishing Rabo-de-Bacalhau apartments were formulated. Second, each assignment was addressed by the RB-grammar and by 4 professional architects. Third, the RB-grammar solution and the 4 architect designs were for each assignment ranked by design quality by architectural evaluators.

5.1 The assignments

The 3 assignments were labelled A, B and C, and consisted each of an existing layout of a Rabo-de-Bacalhau apartment (Figure 3) and of a functional programme and brief description of the building construction system (Table 1). The functional programme represents the clients' wishes, say of the family living in the apartment or of the owning company. The architects were informed that the client has a budget that allows some structural adaptions and changes to the layout of kitchen and bathrooms, but that this budget is not unlimited.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

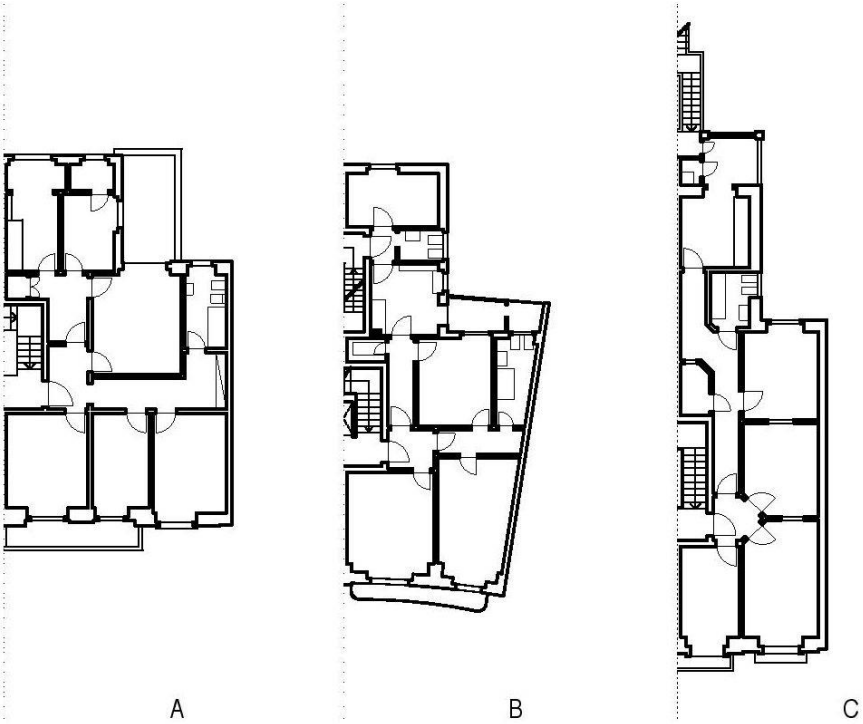

Figure 3: The existing layout of the apartments for assignments A, B and C.

Table 1: The functional programme and description of the building's construction system for assignments A, B and C.

Family A	Couple with 2 children: a boy aged 9 a girl aged 7	**Minimal functional programme**
		1 Double bedroom; 1 Twin bedroom; 1 Kitchen; 1 Living room; 1 Bathroom (toilet, lavatory, bidet, bath or shower)
		Extra areas or functions in order of priority
		2 Individual bedrooms for the kids; Living room and dining room nearby; 2 Private bathrooms; 1 Guest bathroom; Office space integrated in one room or independent
		Important connections between rooms
		Children's bedrooms near parents' bedroom; Private bathrooms in private area; Dining room near kitchen; Living room near entrance
Family B	Couple without children	**Minimal functional programme**
		1 Double bedroom; 1 Kitchen; 1 Living room; 1 Bathroom (toilet, lavatory, bidet, bath or shower)
		Extra areas or functions in order of priority
		Independent office space; Large living/dining room ($\geq 25m^2$); 1 Guest bathroom
		Important connections between rooms
		Private areas (bedroom and private bathroom) segregated from the other areas
Family C	Elderly and young adult	**Minimal functional programme**
		2 Individual bedrooms; 1 Kitchen; 1 Living room; 1 Bathroom (toilet, lavatory, bidet, bath or shower)
		Extra areas or functions in order of priority
		Eating area in the kitchen; Living/dining room or living room and dining room connected; 1 Guest bathroom
		Important connections between rooms
		Dining room near kitchen
Apartments A, B and C		Building built in the 1940s/1950s with a mixed structure of reinforced concrete and stone or brick masonry. Structures have thin concrete slabs and a few pillars and beams. Interior walls are made of brick masonry and several are loadbearing.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

5.2 The architects

There were 6 architects requested to participate in the experiment. These were selected to represent the group of architects that could become involved in refurbishing Rabo-de-Bacalhau apartments, i.e., of professionally practicing architects familiar with Portuguese architecture, including Rabo-de-Bacalhau architecture. First, we selected from the faculty members of the Department of Architecture and Urbanism of the ISCTE University Institute of Lisbon (ISCTE-IUL) all architects who fulfilled these criteria and were not part of the group of evaluators (see next subsection), or had any prior knowledge of the experiment. This led to 3 architects, of which 2 had their training in the Porto Architectural School and 1 in the Lisbon Architectural School. For increasing the representation of the Lisbon School we added architects with a Lisbon background. So, second, we introduced from the ISCTE-IUL network a fourth architect with such a background. The resulting group of 4 architects was dominated by mid-career architects with working experience ranging from 16 to 25 years. We therefore added, thirdly, two more architects: a junior with 4 years of experience, and a senior with 38 years. The work the architects had to do in the experiment was pro bono, which limited possibilities to select more randomly architects.

The architects were labelled I, II, … VI, and for limiting workload each architect was presented 2 of the 3 assignments.

5.3 The evaluators

The evaluators were selected from faculty members of the Department of Architecture and Urbanism of ISCTE-IUL. Selection criteria were that they were trained as architects and formally and practically engaged in the supervision and guidance of architectural master projects at the Department in the academic year of 2013/2014. All faculty members fitting these criteria were invited, which led to 12 evaluators, labelled 1, 2, … 12. For limiting workload each evaluator ranked 2 of the 3 assignments (see Table 11 in the Appendix).

5.4 The experimental protocol

The assignments were evenly and randomly distributed over the architects. This led to 4 architect designs for each of the 3 assignments.

The experimental protocol encompassed three parts. The first and second parts consisted of oral interactions with the architects, and the third of a written interaction. Since the task requested from the architects was pro bono yet time consuming, the architects were addressed personally to ensure their acceptance and guarantee to accomplish it. After this first part the authors orally explained the design tasks to the architects in a uniform manner, not mentioning the experiment's aim. The third part consisted of delivering to the architects, in a closed envelope, the drawings of the existing layouts of the apartments and the written descriptions of the task. The architects returned their designs by hand or by email.

For each of the assignments also a solution was generated with the RB-grammar. This grammar is not yet an implemented tool, hence the authors generated the solution by applying the RB-grammar algorithm to the assignments which leads to a good approximation of the RB-grammar solution; a more ideal experiment would not consist of letting other operators generate the solutions but of using the implemented system. The grammar solution for each of

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

the assignments is the one generated with the standard parameterization values, i.e., the first option for transformation.

All 5 designs generated for an assignment – the RB-grammar solution plus 4 architect designs – were redrawn in a uniform style as given in Figure 4 to Figure 6. Also the labels used for the designs did not reveal their origin (in the Figure 4 to Figure 6 the RB-grammar solutions are however labelled with a "g" for informing readers which designs are RB-grammar solutions).

Figure 4: Architect designs (I, II, III, IV) and RB-grammar solution (g) for assignment A.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

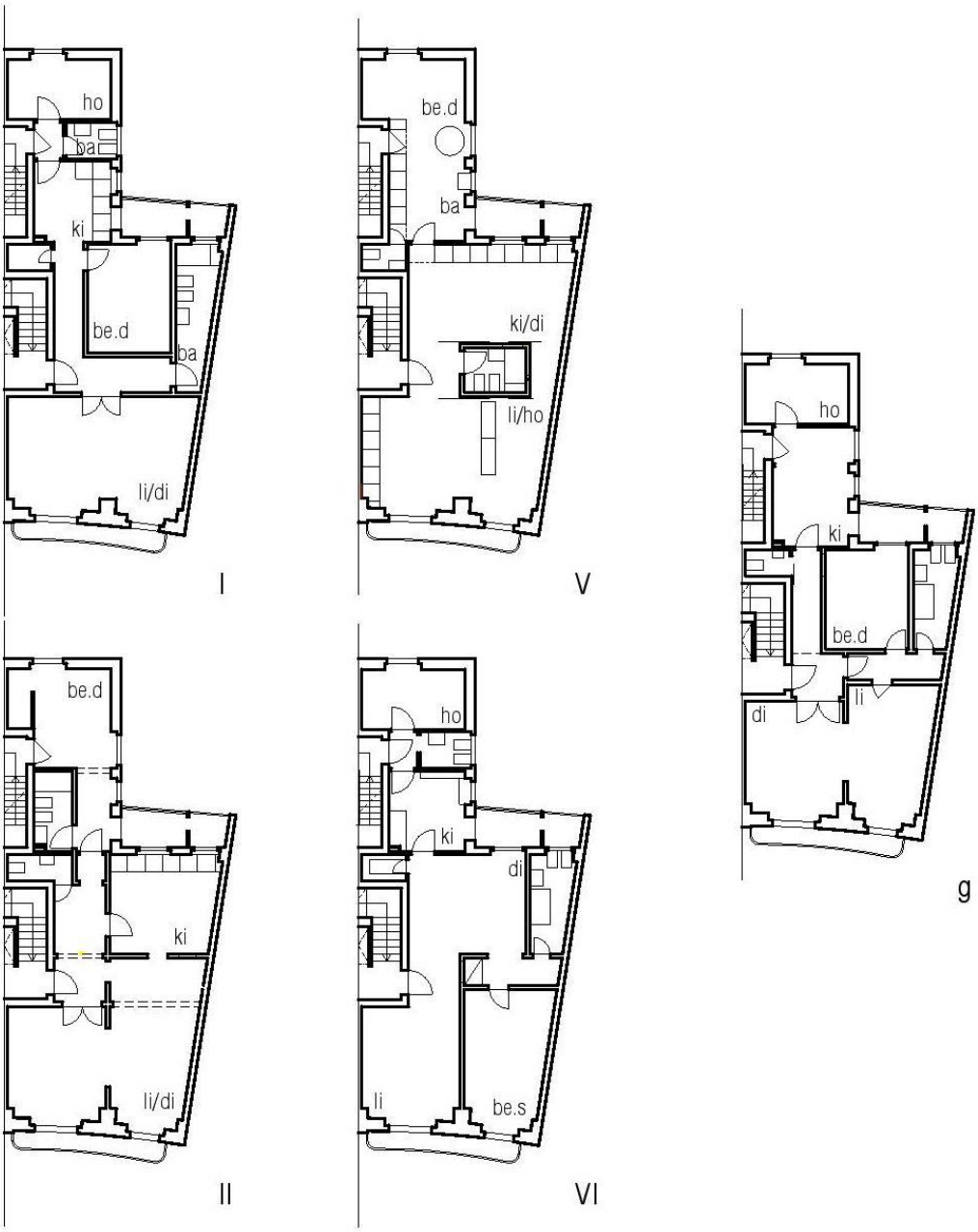

Figure 5: Architect designs (I, II, IV, VI) and RB-grammar solution (g) for assignment B.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

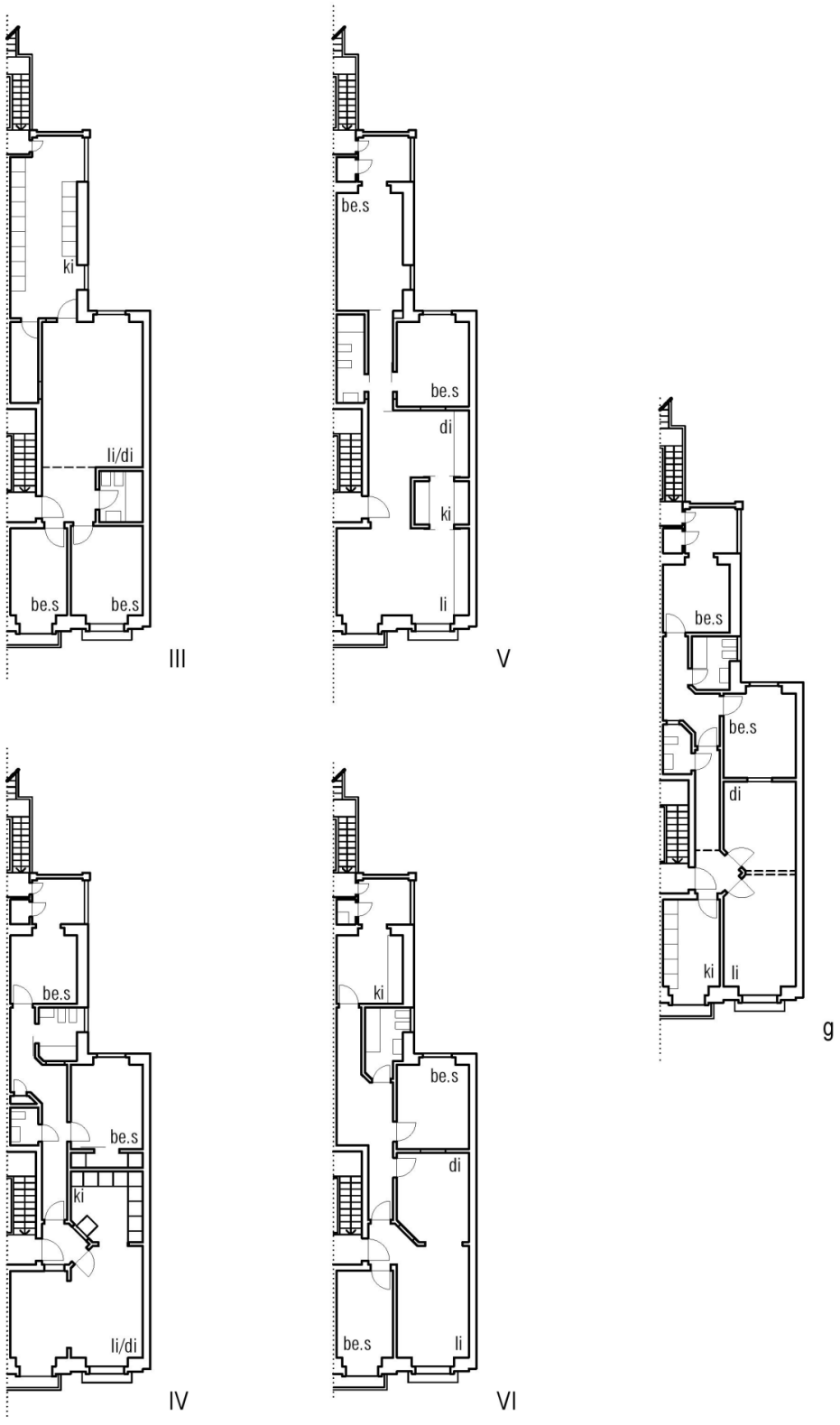

Figure 6: Architect designs (III, IV, V, VI) and RB-grammar solution (g) for assignment C.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

To each evaluator the outcomes for 2 assignments were presented, say, the 5 designs for assignment A and the 5 for assignment C. These assignments were randomly distributed over the evaluators. The evaluators were asked to rank for each assignment the 5 designs by their quality, with the design with highest quality first and the design with the lowest quality last. The evaluators were neither informed about the experiment's aim nor about that grammar solutions were part of the sets of 5 designs. The designs were presented in paper, each printed on a different sheet with the final floor plan (printed in black) and a floor plan highlighting the demolishing elements (yellow) and new constructions (red). The assessments were returned both in paper and by email. The evaluators received the ranking request in written form and could get upfront oral explanation from the authors. Yet, during the ranking of the designs, evaluators could not interact with the authors. This enabled us to retrieve the quality rankings independently of all kinds of additional comments or suggestions by evaluators, say that a specific design was actually quite good if some changes were made.

The written request to the evaluators reads:[3]

> The client – a family that bought a *Rabo-de-Bacalhau* apartment – has commissioned five architects to come up with plans for refurbishing his/her apartment. The family has given a list of their minimal requirements, of their additional wishes, and has indicated that they have a budget that allows them to support, e.g., some structural adaptions related to layout alterations as well as relocating kitchens and bathrooms. Still their budget is not unlimited.
>
> The client has now received five preliminary designs for the general layout of the apartment.
>
> Can you give a recommendation to the client which proposal to choose by ordering the five preliminary designs by design quality? The preliminary designs should be considered as finished; the architect of the chosen design will detail his or her design but will not change the general layout. After that phase of detailing, the design is sent away to the commissioned construction firm, and the works start.

The outcomes of the experiment are given in Section 6 after which they are analysed. In Sections 8 and 9 this analysis is done with statistical tools and we then use a significance level of 0.05. This means that a result is considered statistically significant when the p-value is less than this significance level of 0.05, where the p-value is the probability of rejecting a null hypothesis given that it is true (these null hypotheses are given in Sections 8 and 9).

5.5 Bias and other shortcomings

The presented experiment can be criticised on a number of points.

The 3 assignments were not randomly chosen but defined by the authors: the functional programmes are posited as paradigmatic for inhabitants of Rabo-de-Bacalhau apartment, and the apartments' layouts were chosen considering those inhabitants and the Rabo-de-Bacalhau types according to the method referred in (Eloy, 2012).

[3] The experiment was conducted in Portuguese.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

The 6 architects were not randomly chosen but drawn from the network of the Department of Architecture and Urbanism of ISCTE-IUL. These architects knew that the assignments would also be carried out by other architects, but they neither knew who the others architects were nor that a shape grammar was applied to the assignments.

The evaluators are representative for ISCTE-IUL, since all department members meeting the criteria defined in Section 5.3 were invited to participate in the experiment. The choice of ISCTE-IUL is not random. The evaluators did the evaluation in their own offices and were asked not to talk about the designs with colleagues.

6 The outcomes

With 10 of the 12 evaluators responding, the experiment produced 6 design rankings for assignment A, 7 for B and 7 for C (Table 2; in the Appendix evaluators and rankings are correlated, Table 11).

Table 2: Quality rankings of the architect designs and RB-grammar solutions; the design judged to be of highest quality comes first.

Assignment A	Assignment B	Assignment C
g; I; III; IV; II	g; I; II, VI, V	g; IV; III; V; VI
IV; g; I; III, II	I; VI; g; V; II	g; VI; III; IV; V
III; II; IV; g; I	II; V; g; I; VI	g; IV; VI; V, III
I; g; III; IV; II	g; I; VI; II; V	g; III; VI; IV; V
g; I; IV; III; II	VI; I; g; V; II	III; IV; g; VI, V
g; I, IV, II, III	VI; I; II; g; V	V; III; g; IV; VI
	VI; g; I; II; V	VI; g; III; IV; V

For the analysis of these quality rankings we introduce the numerical value r(g) as the rank the RB-grammar solution g gets in a specific ranking, and r(X) as the rank a specific architect design X gets in that ranking. For each ranking the values of r(g) and r(X) run from 1 to 5; for the top left ranking in Table 2, for instance, r(g) is equal to 1, and r(I) equal to 2. The outcomes of the experiment can then be summarised in terms of the distribution of the ranks r(g) the RB-grammar solutions received (see Table 3).

Table 3: The distribution of the RB-grammar solution rank r(g) per assignment and overall.

	i	1	2	3	4	5
assignment A	# r(g) = i, out of 6	*3*	*2*	*0*	*1*	*0*
assignment B	# r(g) = i, out of 7	*2*	*1*	*3*	*1*	*0*
assignment C	# r(g) = i, out of 7	*4*	*1*	*2*	*0*	*0*
all assignments	# r(g) = i, out of 20	*9*	*4*	*5*	*2*	*0*

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

7 Aggregated quality rankings

The goal of the experiment is to compare the design quality of the RB-grammar solutions with the design quality of the architect designs. Our criterion for taking the quality of an RB-grammar solution as similar to the quality of the architect designs in a ranking is that the RB-grammar solution comes third, since then there are two architect designs with a quality higher than the RB-grammar solution, and two architect designs with a quality lower than the RB-grammar solution.

With this criterion and the found quality rankings (Table 2) it can be concluded that the majority of evaluators ranked the RB-grammar solutions as being of similar or higher quality in comparison with the designs of the professional architects (see Table 3): for assignment A the RB-grammar solution was evaluated 5 out of 6 times as being first, second or third in quality rank; for B it was 6 times out of 7; and for C 7 out of 7.

For assessing the design quality of the RB-grammar solutions for each assignment, the found rankings in Table 2 should be aggregated per assignment. Let R(g) be the aggregated rank of the RB-grammar solution g, and let R(X) be the aggregated rank of the architect design X. This aggregation is not unambiguous, yet two ways come rapidly to mind using the ranks r(g) and r(X) as defined by each quality ranking in Table 2. First, one can take the ranks r(g) and r(X) as defining an ordinal scale. The best overall design with aggregated rank R(X)=1 is then the design that comes first most often (and in case of *ex aequo*, one looks at the number of times the design comes second, et cetera). Second, one can take the ranks r(g) and r(X) as defining an interval scale, over which numerical averages can be taken. A design X that gets a 2 in one ranking and a 5 in a second, then has for both rankings an aggregated rank R(X) of 3.5. When one does the aggregation using these two ways and again ranks the different designs per assignment, one obtains qualitatively a fairly consistent outcome (Table 4). For assignments A and C the RB-grammar solution has an aggregated rank R(g) of 1, meaning that overall the RB-grammar produced the best solution. For assignment B the aggregated rank R(g) is 2, meaning that overall the RB-grammar produced the second best solution.

Table 4: Aggregated quality rankings per assignment.

	Assignment A	Assignment B	Assignment C
ordered by highest rank	g; I; IV; III; II[4]	VI; g; I; II; V	g; III; VI; V; IV
ordered by average rank	g; I; IV; III; II[5]	I; g/VI; II; V[6]	g; III; IV; VI; V

When applied to the rankings aggregated per assignment, our criterion for taking an RB-grammar solution as having a design quality similar to the architect designs is that the aggregated rank R(g) of the grammar solution is 3. A result of the experiment is thus that the design quality of the RB-grammar solutions is higher than a quality similar to the architect designs.

[4] The RB-grammar solution comes first for assignment A because g is most often ranked as best; the designs I, III and IV are all one time ranked as best and score with respect to that equal, yet design I was three times ranked as second best; design IV three times as third best, and design III two times as third best.
[5] The average rank values for assignment A are: R(g)=1.83; R(I)=2.50; R(IV)=3.00; R(III)=3.33; R(II)=4.33.
[6] For assignment B the average ranks R(g) and R(VI) are both 2.43.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

For understanding this first result of the experiment, the performances of the architects relative to the RB-grammar may be related to their designs and their professional characteristics. Figure 7 shows the individual performances of the architects and of the RB-grammar, overall and per assignment (taking the ranks r(g) and r(X) as defining an interval scale).

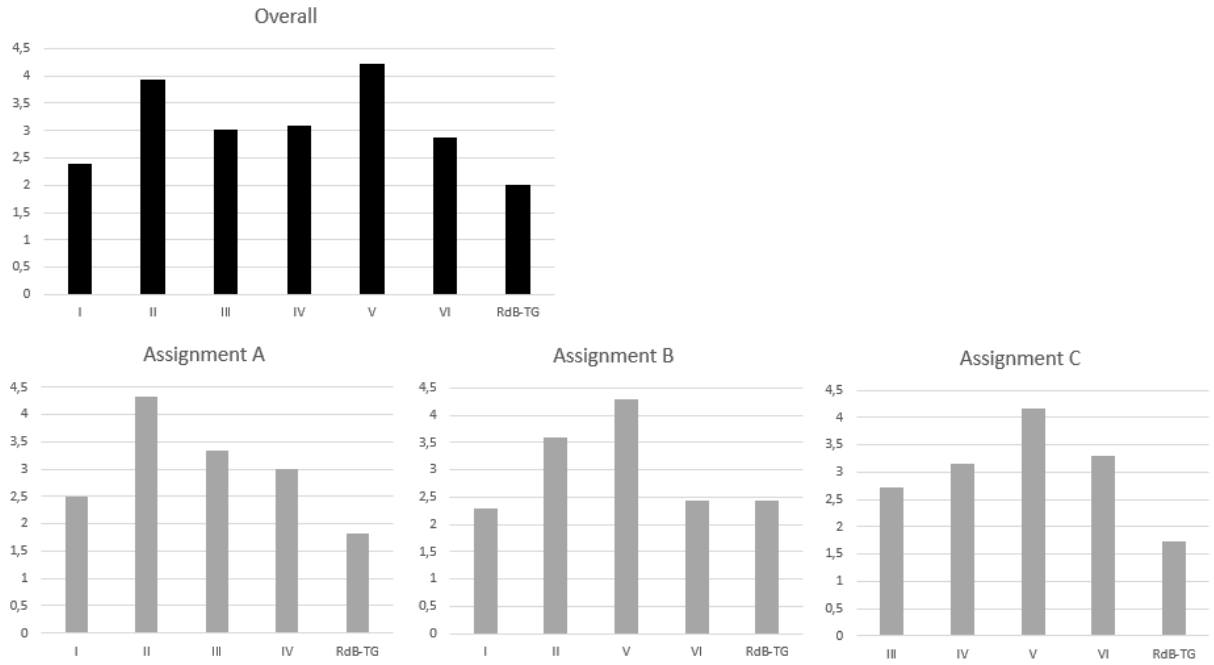

Figure 7: The overall averages (top) and averages per assignment for each architect and RB-grammar.

The overall averages show that the designs by the RB-grammar and of Architect I were ranked as of best quality. These designs were typically not very intrusive to the original layout of the apartments and addressed the majority of the clients' requirements. Architect I is a mid-career architect with several years of practice and well experienced in dealing with functional programmes both in new constructions and refurbishment work. The good results of the RB-grammar may be explained by the fact that its transformation rules for architectural elements incorporate a vast amount of design knowledge about Rabo-de-Bacalhau apartments (Eloy, 2012). Therefore one may say that the RB-grammar is the best informed designer about Rabo-de-Bacalhau apartments compared to any individual architect.[7]

The designs by Architects III, IV and VI form a group that were ranked as of lesser quality, and these architects have rather different characteristics. Architect III made more intrusive proposals but addressed the clients' requirements, Architect IV is the oldest architect with a vast experience in refurbishment, and Architect VI is the youngest and less experienced architect. The design for assignment B by Architect VI was considered to be the best by three evaluators. This outcome may be explained by his option for small interventions, which seem to be a characteristic appreciated by the evaluators.

[7] We are grateful to one of the reviewers for pointing out this perspective on shape grammars.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

The architects whose designs were ranked of lowest quality, II and V, are similar in terms of years of experience and age. Architect II and specially Architect V proposed designs that require major demolitions and constructions and are therefore not respecting the original layout of the Rabo-de-Bacalhau apartments.

The preference of the evaluators for less intrusive designs may be related to the clause in the assignments that the clients had a limited budget. Or it can be related to the current practice in refurbishment to value the maintenance of the identity of the existing buildings with respect to construction and functional characteristics.

8 Statistical quality ranking

The analysis of the experiment in the previous section does not accommodate the possibility that the quality of the designs was subject to random fluctuations, i.e., that an RB-grammar solution was accidently of high quality or that an architect had a bad day and created a low quality design. This possibility can be incorporated by analysing the found rankings statistically.

Statistical analysis means introducing a null hypothesis H_0 about the rankings and determining whether the experiment's outcomes give evidence to reject it in favour of an alternative hypothesis H_1. Consider the case that an evaluator compares the design quality of one design by a specific architect with the quality of a RB-grammar solution. Let H_0 be that there is then a high probability p that the evaluator takes the quality of the architect design as higher than that of the RB-grammar solution; it will then not often occur that an RB-grammar solution is accidently very good, or that an architect has a bad day. Let H_1 be that this probability p is not high:

H_0: The probability that an evaluator takes the quality of an architect design as higher than the quality of an RB-grammar solution is p.

vs

H_1: The probability that an evaluator takes the quality of an architect design as higher than the quality of an RB-grammar solution is smaller than p.

If the experiment yields a ranking r(g), then an evaluator decided r(g)-1 times out of 4 that the quality of an architect design is lower than the quality of the RB-grammar solution. The probability distribution (Table 5) that the null hypothesis H_0 gives for r(g) is then equal to the binominal distribution Bi(n,p) with n=4 (assuming that the design qualities of the four architect designs are independent of each other):

Table 5: Theoretical probability distribution of r(g) under H_0.

I	1	2	3	4	5
P(r(g)=i)	$(1-p)^4$	$4\,p(1-p)^3$	$6\,p^2(1-p)^2$	$4\,p^3(1-p)$	p^4

With p having a high value, the probability that r(g) is 4 or 5 in an individual ranking becomes high as well. Yet occasionally a lower rank (r(g) is 1, 2 or 3) may turn up as well. Let us accept that H_0 allows that this latter possibility may occur in about one out of four rankings.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

This acceptance means choosing p in H_0 as equal to 0.75, since then the probability that r(g) is 1, 2 or 3 is equal to 0.26 (see Table 6).

Table 6: Theoretical probability distribution of r(g) under H_0 with p=0.75.

I	1	2	3	4	5
P(r(g)=i)	0.00	0.05	0.21	0.42	0.32

When the distributions of the rankings r(g) obtained in the experiment are compared to the probability distribution that H_0 predicts (Table 6), there seems to be a mismatch (Figure 8). To determine whether this mismatch is statistically significant one can do a chi-square test.

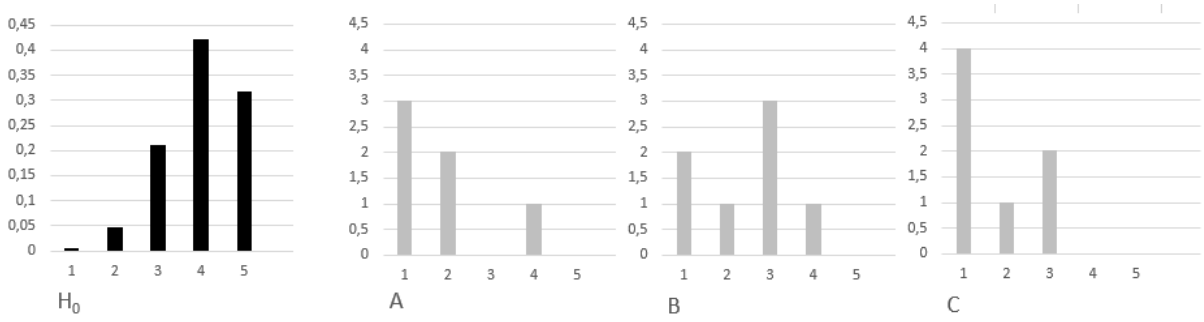

Figure 8: Probability distribution of r(g) under H_0 for p=0.75 (left) and distributions of the outcomes for r(g) per assignment.

When this chi-square test is carried out rigorously it should be conducted for the three assignments A, B and C separately in order to avoid mixing data of what in principle are different cases. Yet in order to meet the conditions of a chi-square test (one of them is that the expectation value for an outcome is 5 or higher), we consider all rankings on a par. For this same reason we consider a coarse-grained description of the outcomes of the experiment, where the rank r(g) of the RB-grammar solution can take three values: "high", 4 and 5 (so the outcome "high" lumps together the outcomes that r(g) is 1, 2 or 3). The predictions of H_0 then simplify to a fairly even probability distribution which can be tested against the found numbers of outcomes (see Table 7).

Table 7: Coarse-grained probability distribution under H_0, and expected and found outcomes of the experiment.

r(g)	high	4	5
Probability distribution under H_0	0.26	0.42	0.32
Expected numbers of outcomes out of 20	5.23	8.44	6.33
Found numbers of outcomes	18	2	0

A chi-square test applied to these expected and found numbers of outcomes gives a result of 42.377 and leads to a one-tailed p value close to zero, which is substantially smaller than the significance level of 0.05, such that we can strongly reject H_0 in favour of the alternative H_1. Hence, an evaluator assesses an architect design as of higher quality than an RB-grammar solution with a probability lower than 0.75.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

9 Quality rank distributions

With the above analyses one can conclude that the RB-grammar has produced in the experiment solutions that have a similar or higher quality as the designs created by the architects, even if that quality is subject to random fluctuations. This conclusion does not yet imply that the design quality of the RB-grammar solutions is in all respects similar to the quality of the architect designs. For instance, if one focusses on the design quality rank *distributions* that the experiment generated for the RB-grammar solutions and the architect designs, there seem to be profound differences (see Figure 9 for the frequencies of ranks r(X) and r(g) when no distinction is made between the different architects and different assignments). Hence, our outcomes seem also to prove that design quality the RB-grammar produces is not similar to that of professional architects. This first impression can however be qualified: further statistical analysis shows that the quality rank distribution of the RB-grammar solutions is in terms of the mean rank, similar to the quality rank distributions of the designs of two of the six architects. So, with respect to this mean rank, the design quality of the RB-grammar is similar to the design quality of *some* professional architects.

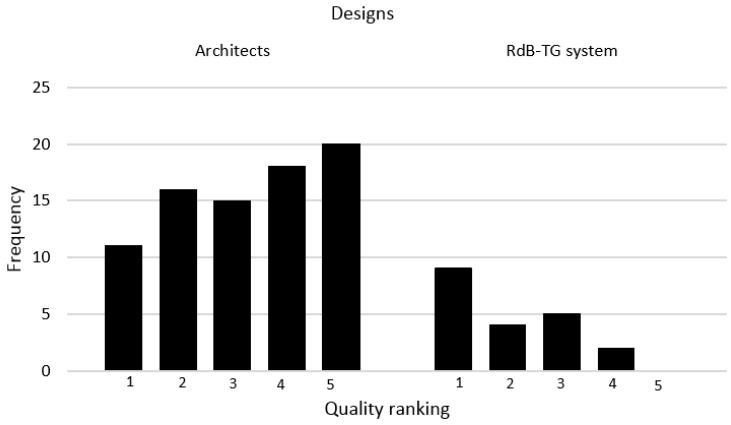

Figure 9: Frequencies of the quality ranks of the architect designs and the RB-grammar solutions.

The first step in our further statistical analysis is comparing the quality rank distributions of the architect designs and the RB-grammar solutions. The competing hypotheses are now:

H_0: The population of the architect designs and the population of the RB-grammar solutions are equal with respect to their quality rank distributions.

vs

H_1: The population of the architect designs and the population of the RB-grammar solutions are not equal with respect to their quality rank distributions.

The experiment produced a smaller number of outcomes, either for the three assignments individually or collectively. Together with the fact that these outcomes are ranks, we thus cannot assume a normal distribution to the variable that ranks the designs. A nonparametric analysis for assessing H_0 is therefore most adequate. A nonparametric test to compare outcomes between two independent groups is the Mann-Whitney test for determining whether two samples are likely to derive from the same population. The results of this test as applied to the samples as given in Figure 9 are presented in Table 8, and as can be observed, the null hypothesis H_0 is rejected by these results for any level of significance. Hence, there is

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

significant evidence that the architect designs and the RB-grammar solutions have different quality rank distributions.

Table 8: Mann-Whitney test results to determine if the architect designs and RB-grammar solutions derive from the same population.

	Designs	
	Architects	RB-grammar
Mean Rank	55.5	30.5
n_j	80	20
Sum of Ranks	4440.0	610.0

	Value	Significance level
W	610.0	0.00

This result does not yet reveal what precisely the differences are between the quality rank distributions of the architect designs and RB-grammar solutions, and ignores the possibility that this difference may not be present between the designs by specific individual architects and RB-grammar. In a second step we therefore analyse the quality rank distributions per individual architect and RB-grammar, focussing on their mean ranks.

The quality rank distributions of the architect designs split out per individual architect (still not making distinctions between the different assignments) again show obvious differences with those of the RB-grammar solutions (Figure 10).

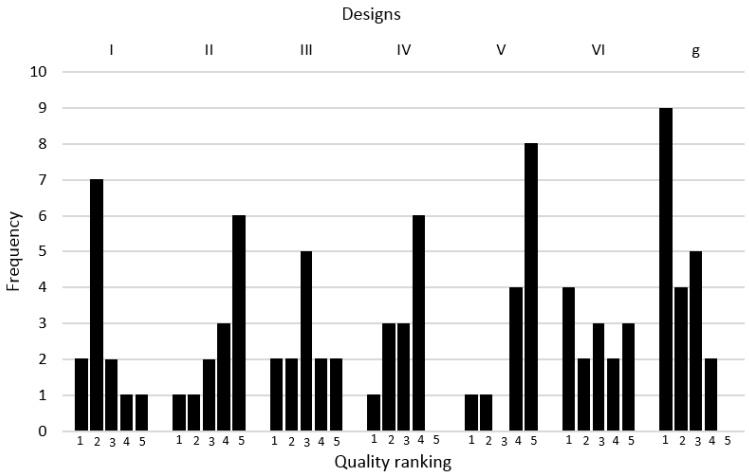

Figure 10: Frequencies of the quality ranks of the designs per architect and of the RB-grammar solutions.

For testing whether the mean ranks of these 7 distributions are equal, we use the Kruskal-Wallis test for k comparison groups (k>2), which is sometimes described as an ANOVA[8] with the data replaced by their ranks, here k=7. The null and alternative hypotheses for the Kruskal-Wallis nonparametric test are:

[8] ANOVA provides a parametric test for comparing multiple means. It requires that a number of conditions are met: independent samples; equal variances between samples and all the populations must be normally distributed.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

H_0: The architect designs and RB-grammar solutions are equal with respect to the mean ranks of their quality rank distributions.

vs

H_1: At least 1 of the architect designs or RB-grammar solutions is different with respect to the mean ranks of their quality rank distributions.

The results are summarised in Table 9. The test comparisons of the mean ranks proceeds by pooling the observations from the 7 samples into one combined sample, keeping track of which sample each observations comes from, and then ranking lowest to highest from 1 to N, with $N=n_1+n_2+\ldots+n_7$ (here N=100).

Table 9: Kruskal-Wallis test results to compare if architect designs and RB-grammar solutions are equal with respect the mean ranks of their rank quality distributions.

	Designs						
	I	II	III	IV	V	VI	g
Mean Rank	38.19	68.96	50.50	52.04	74.79	47.64	30.50
n_j	13	13	13	13	14	14	20

	Value	Degrees of freedom	Significance Assymp. Sig.)
Chi-Square	28.22	6	0.00

The null hypothesis H_0 is rejected for any value of significance: there is statistically significant evidence that there is a difference in the mean ranks of the quality rank distributions of the designs.

The Kruskal-Wallis Test does not specify which of the quality rank distributions are the different ones with respect to the mean rank. A third and final analysis step is therefore a multiple comparison of the mean ranks of the different quality rank distributions. When comparing these mean ranks 2 by 2 using the Least Significant Distance (LSD) Fisher test, it becomes clear that the RB-grammar solutions exhibit differences with the designs by Architects II, III, IV and V, and not with the designs of Architects I and VI (see Table 10; the * sign indicates that the mean difference is significant at the 0.05 level). As it can be noticed the difference between the architect designs and g (i-j) is positive reinforcing that the mean ranks of the quality rank distributions for the RB-grammar solutions are always smaller than the mean ranks of the quality rank distributions for the architect designs, which according to the authors' position expresses the "good" results obtained for the RB-grammar solutions as compared to the architect designs.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

Table 10: Least Significant Difference test results for the differences of the mean ranks of the quality rank distributions.

Designs	Mean Difference (i-j) LSD						
(i) (j)	I	II	III	IV	V	VI	g
I		-30.77*	-12.31	-13.85	-36.59*	-9.45	7.69
II			18.46	16.92	-5.82	21.32*	38.46*
III				-1.54	-24.29*	2.86	20.00*
IV					-22.47*	4.40	21.54*
V						27.14*	44.29*
VI							17.14
G							

The same procedure was carried out but now considering the designs for each assignment separately. The number of observations are lower (6 or 7) and the differences are less pronounced, but for each assignment there is at least one architect design with a statistically significant different mean rank (see the Appendix for the Kruskal-Wallis test). The multiple comparisons of the mean ranks of the quality rank distributions for assignment A makes clear that design II is different from the remaining ones, for assignment B design V is notoriously different from the others, as it is for assignment C. However it must be remembered that the designs for the three assignments were done by different sets of architects, hence formally a comparison between these assignments is not possible.

The upshot of this final analysis is that when the designs are considered all together, then with respect to the mean rank of the quality rank distributions the architect designs and the RB-grammar solutions can be ordered in two groups: designs I, VI and g in one group and designs II, III, IV and V in another. This supports the conclusion that, with respect to this mean rank, the design quality as generated by the RB-grammar is similar to the design quality of at least two of the six professional architects.

For understanding this final result of the experiment, one can again consider the different designs and the professional characteristics of the architects. The designs by the RB-grammar, Architects I and VI, making up the first group with lower mean ranks of their quality rank distributions, were the ones that more persistently proposed less intrusive interventions and thus maintained the characteristics of the Rabo-de-Bacalhau apartments. Architect I is, as said in Section 7, a senior architect with several years of experience. Architect VI is the youngest architect who proposed simple and non-intrusive layouts. And the RB-grammar arrived at its solutions with transformation rules for architectural elements informed by design knowledge about Rabo-de-Bacalhau apartments.

The designs of Architects II, III, IV and V make up the second group of designs with higher mean ranks of their quality rank distributions. This lower quality performance may be due to the more intrusive strategies the architects in this second group followed in comparison with the first group. The architects in this group approached the refurbishment more freely, as if it concerned new construction. For some of these architects the choice for these strategies may be related to their having less experience with refurbishment.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

10 Conclusions and discussion

In this paper an experiment was presented for determining whether shape grammar systems can solve architectural tasks with a design quality similar to that of professional architects. The architectural task we considered was the refurbishment of Lisbon's Rabo-de-Bacalhau apartments, and the shape grammar we used was the Rabo-de-Bacalhau transformation grammar (RB-grammar). Analysis of the outcomes of this experiment gave three results. First, a majority architectural evaluators judged that the RB-grammar generated solutions with an (aggregated) design quality that is higher than the quality of refurbishment designs by professional architects. Second, this result about the design quality of shape grammar systems holds even if it is assumed that the quality of the RB-grammar and of the architect designs is subject to random fluctuation. Third, when one considers the mean rank of the quality distributions, then the RB-grammar is indistinguishable from at least two of the six professional architects participating in the experiment.

In the experiment the solutions of the RB-grammar were compared with preliminary designs by the architects, and we argued that this comparison makes architecturally sense for the task of refurbishing the full stock of Rabo-de-Bacalhau apartments in the city of Lisbon. This comparison may however be criticised. It may, for instance, be noted that the RB-grammar incorporates rules to maintain the structural stability of the buildings, whereas architects could consider this stability not in their preliminary design but in later phases of their design processes. It therefore makes sense to also consider experiments that compare the solutions created by shape grammar systems with more detailed designs by professional architects.

When accepting the argument that Lisbon's full stock of Rabo-de-Bacalhau apartments can be refurbished either by using the RB-grammar or by commissioning preliminary designs by professional architects, the conclusion of this paper is that the first option may lead to similar architectural design quality as the second, or even to higher quality. This conclusion suggests further research about how the RB-grammar can be made available to the owners of Rabo-de-Bacalhau apartments. Questions that emerge are then what kind of interface is needed? Who will use it, how will they used it, and what comes after the definition of the final layout by the RB-grammar? What problems may emerge if owners adopt designs of the RB-grammar for refurbishing their apartments, and where lies the responsibility if adopted designs turn out to be faulty? (Eloy and Vermaas, 2014)

The case that we used in this paper for our arguments is the case of the RB-grammar and the refurbishment of Lisbon's Rabo-de-Bacalhau apartments. A third line of further research is to determine if also other architectural tasks can be addressed by shape grammar systems. There are, for instance, many other large stocks of types of multifamily housing buildings in cities around the world that, as Rabo-de-Bacalhau buildings, are recognisable by their common morphological characteristics. If these buildings are similarly in need of refurbishment, shape grammar systems may play a role in their refurbishment as well. The question then becomes under what conditions it makes practically sense to develop such systems for architectural tasks. One such condition seems to be that such an architectural task is highly repetitive. The value of shape grammar systems lies in large-scale, customized and standardised architectural tasks; for addressing more singular tasks human architects are to be preferred for reasons of economics and creativity. A second condition may be that the knowledge needed to properly address an architectural task is hard and time consuming to collect. This knowledge may therefore often not fully be applied by human architects, whereas shape grammar systems are

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

developed on the basis of all this knowledge and thus incorporate it in their rules. The RB-grammar, for instance, was created after a large investigation of Rabo-de-Bacalhau buildings and their refurbishment, hence uses existing knowledge about this refurbishment more completely than human architects.

At the end of this line of research shape grammar systems may emerge as valuable assets to architecture for taking up large-scale refurbishment projects. Where such projects are currently practically or economically unfeasible for professional architects due to the vastness of the effort to redesign thousands of apartments on an individual basis, large-scale refurbishment can become available to architecture when shape grammar systems are taken as stand-alone design systems.

11 Acknowledgements

We are grateful to the designers for doing the professional architectural designs for the experiment and to the evaluators for ranking the designs. We thank Elisângela Vilar and especially the reviewers for helpful suggestions.

12 References

Abellem K (1979) *Système d'evaluation de logements*. Bern: Office Féderal du Logement.

Arthurson K (2012) *Social mix and the city: challenging the mixed communities consensus in housing and urban planning policies*. Collingwood: Csiro Publishing.

Association Qualitel (1989) *Guide qualitel*. Paris: Association Qualitel.

Eloy S, Vermaas P (2014) Towards Effective City Rejuvenation with ICT: web-based shape grammar supported refurbishment design. In Zreik K (ed) *Architecture, City & Information Design*, pp.129-139.

Cabrita R, Freitas MJ, Pedro JB (1998) Understanding housing satisfaction. In V Arantes, O Ural (Eds.), *Proceedings of the XXV IAHS World Housing Congress*. Vila Nova de Gaia: Faculdade de Engenharia da Universidade do Porto, pp. 1-9.

CML (2007) *Plano estratégico nacional para a política de habitação: síntese do fórum de Lisboa*. Lisboa: Câmara Municipal de Lisboa.

Colakoglu B (2005) Design by grammar: hybrid applications of grammar in design studio projects. In *Predicting the future, 25th eCAADe conference proceedings*. Frankfurt am Main, pp. 919-925.

Duarte JP, Beirão J (2011) Towards a methodology for flexible urban design: designing with urban patterns and shape grammars. *Environment and Planning B: Planning and Design* 38(5): 879-902.

Eloy S (2012) *A transformation grammar-based methodology for housing rehabilitation.* PhD thesis. Lisboa: Instituto Superior Técnico, Universidade Técnica de Lisboa.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

Eloy S, Duarte JP (2014) Inferring a shape grammar: translating designer's knowledge. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 28(2): 153-168.

Flemming U (1994) Get with the program: common fallacies in critiques of computer-aided architectural design. *Environment and Planning B: Planning and Design* 21(7): 106-116.

Hall K, Warm P (1994) *Greener building: products and services directory*. Coaley: Association of Environment Conscious Building.

INE (2012) *Censos 2011*. Online: Instituto Nacional de Estatística. Availabe at WWW: <http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos2011_apresentacao>

Jacob BL (1996) Algorithmic composition as a model of creativity. *Organised Sound* 1: 157-165.

Lara F, Youngchul K (2010) Built global, lived local: a study of how two diametrically opposed cultures reacted to similar modern housing solutions. *Journal of Architectural and Planning Research* 27(2): 91-106.

Lees L (2008) Gentrification and social mixing: towards an inclusive urban renaissance. *Urban Studies* 45(12): 2449-2470.

Maginn P, Thompson S, Tonts M (2008) *Qualitative housing analysis: an international perspective*. Bingley: Emerald Group Publishing.

March L (1981) A class of grids. *Environment and Planning B* 8(3): 325-332.

McDermott J (2008) *Evolutionary computation applied to the control of sound synthesis*. PhD thesis. Limerick: University of Limerick.

Monteys X (2013) O edifício comum, casas lisboetas. In CD Coelho (Ed.) *Os elementos urbanos 1, Cadernos MUrb morfologia urbana estudos da cidade portuguesa*. Lisboa: Argumentum, pp. 189-205.

Natividade-Jesus E, Coutinho-Rodrigues J, Antunes C (2007) A multicriteria decision support system for housing evaluation. *Decision Support Systems* 43(3): 779-790.

Pedro JB (2000) *Definição e avaliação da qualidade arquitectónica habitacional*. Phd Thesis. Porto: Faculdade de Arquitetura, Universidade do Porto.

Reffat RM (2008) Digital architecture and reforming the built environment. *Journal of Architectural and Planning Research* 25(2): 118-129.

Sass L (2005) Wood frame grammar: CAD scripting a wood frame house. In *Proceedings of the 11th International Computer Aided Architectural Design Futures 2005 Conference*, pp. 383-392.

Semes SW (2009) *The future of the past: a conservation ethic for architecture, urbanism, and historic preservation*. New York: W.W. Norton & Co.

Stiny G, Gips J (1971) Shape grammars and the generative specification of painting and sculpture. *Information Processing* 71: 1460-1465.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

Stiny G (1976) Two exercises in formal composition. *Environment and Planning B: Planning and Design* 3(2): 187-210.

Stiny G, Mitchell WJ (1978) The Palladian grammar. *Environment and Planning B* 5(1): 5-18.

Tuncer B, Sener S (2010) Generation of a new urban block for the future city: a bottom-up approach for the future city block. In *FUTURE CITIES, 28th eCAADe conference proceedings*, pp. 179-185.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

13 Appendix

In the experiment 12 evaluators were asked to give quality rankings of the architect designs and RB-grammar solutions for two of the three assignments. In Table 11 these rankings are given per evaluator; Evaluator 1 did not submit rankings and Evaluator 5 was not in a position to deliver.

Table 11: The collected rankings; the design judged to be of highest quality comes first.

	Assignment A	Assignment B	Assignment C
evaluator 1:	-	-	
evaluator 2:	g; I; III; IV; II	g; I; II, VI, V	
evaluator 3:	IV; g; I; III, II	I; VI; g; V; II	
evaluator 4:	III; II; IV; g; I	II; V; g; I; VI	
evaluator 5	-		-
evaluator 6:	I; g; III; IV; II		g; IV; III; V; VI
evaluator 7:	g; I; IV; III; II		g; VI; III; IV; V
evaluator 8:	g; I, IV, II, III		g; IV; VI; V, III
evaluator 9:		g; I; VI; II; V	g; III; VI; IV; V
evaluator 10:		VI; I; g; V; II	III; IV; g; VI, V
evaluator 11:		VI; I; II; g; V	V; III; g; IV; VI
evaluator 12:		VI; g; I; II; V	VI; g; III; IV; V

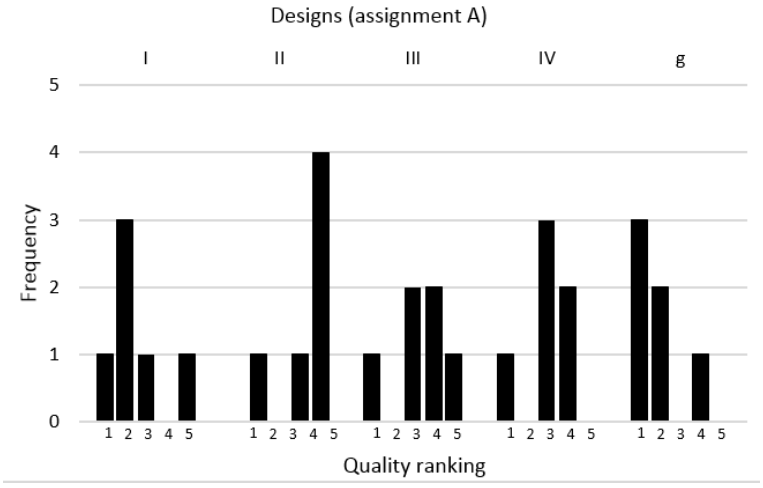

Figure 11: Frequencies of the quality ranks of the designs per architect and of the RB-grammar solutions, for assignment A.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

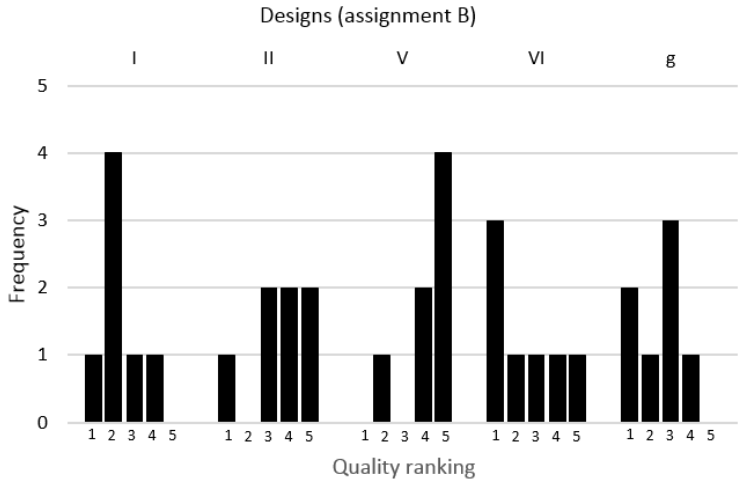

Figure 12: Frequencies of the quality ranks of the designs per architect and of the RB-grammar solutions, for assignment B.

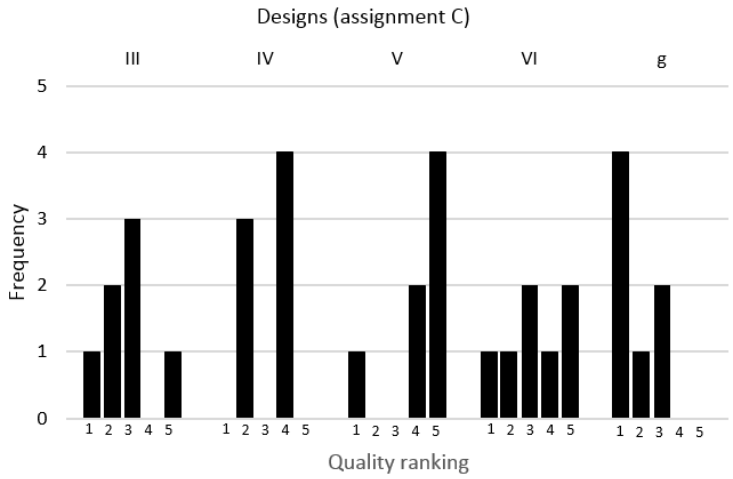

Figure 13: Frequencies of the quality ranks of the designs per architect and of the RB-grammar solutions, for assignment C.

For each assignment W = A, B or C, the hypotheses are:

H_0: For assignment W, the architect designs and RB-grammar solutions are equal with respect to the mean ranks of their quality rank distribution.

vs

H_1: For assignment W, at least 1 of the architect designs or RB-grammar solutions is different with respect to the mean ranks of their quality rank distributions, with respect assignment W.

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.

Table 12: Kruskal-Wallis test results to compare if architect designs and RB-grammar solutions are equal with respect the mean ranks of their rank quality distributions, for assignment A.

	Designs				
	I	II	III	IV	g
Mean Rank	12.5	23.5	17.5	15.5	8.5
n_j	6	6	6	6	6

	Value	Degrees of freedom	Exact Sig.
Chi-Square	10.15	4	0.03

Table 13: Kruskal-Wallis test results to compare if architect designs and RB-grammar solutions are equal with respect the mean ranks of their rank quality distributions, for assignment B.

	Designs				
	I	II	V	VI	g
Mean Rank	13.0	22.0	27.0	14.0	14.0
n_j	7	7	7	7	7

	Value	Degrees of freedom	Exact Sig.
Chi-Square	10.69	4	0.02

Table 14: Kruskal-Wallis test results to compare if architect designs and RB-grammar solutions are equal with respect the mean ranks of their rank quality distributions for assignment C.

	Designs				
	III	IV	V	VI	g
Mean Rank	16.0	19.0	26.0	20.0	9.0
n_j	7	7	7	7	7

	Value	Degrees of freedom	Exact Sig.
Chi-Square	10.69	4	0.02

Pre-print, for citing use: Eloy, S., Vermaas, P.E. & Andrade, M., (2017) The Quality of Designs by Shape Grammar Systems and Architects: A Comparative Test on Refurbishing Lisbon's Rabo-de-Bacalhau Apartments. *Journal of Architecture and Planning Research*. V34, n4, pp.271-294.