# GRL Model Validation: A Statistical Approach

Jameleddine Hassine[1] and Daniel Amyot[2]

[1] College of Computer Science and Engineering, KFUPM, Dhahran, Saudi Arabia
jhassine@kfupm.edu.sa
[2] EECS, University of Ottawa, Ontario, Canada
damyot@eecs.uottawa.ca

**Abstract.** Goal models represent interests, intentions, and strategies of different stakeholders. Reasoning about the goals of a system unavoidably involves the transformation of unclear stakeholder requirements into goal-oriented models. The ability to validate goal models would support the early detection of unclear requirements, ambiguities and conflicts. In this paper, we propose a novel GRL-based validation approach to check the correctness of goal models. Our approach is based on a statistical analysis that helps justify the modeling choices during the construction of the goal model as well as detecting conflicts among the stakeholders of the system. We illustrate our approach using a GRL model for the introduction of a new elective security course in a university.

## 1 Introduction

There is a general consensus on the importance of good Requirements Engineering (RE) approaches for achieving high quality software. Requirements elicitation, modeling, analysis and validation are amongst the main challenges during the development of complex systems. A common starting point in requirements engineering approaches is the elicitation of goals that the targeted system will need to achieve once developed and deployed. Goal modeling can be defined as the activity of representing and reasoning about stakeholder goals using models, in which goals are related through relationships with other goals and/or other model elements, such as, e.g., tasks that system is expected to execute, resources that can be used, or roles that can be played [1]. Over the past two decades, several goal modeling languages have been developed. The most popular ones are *i** [2], the NFR Framework [3], Keep All Objects Satisfied (KAOS) [4], TROPOS [5] and the Goal-oriented Requirement Language (GRL) [6] part of the ITU-T standard User Requirement Notation (URN).

The growing popularity of goal-oriented modeling, and its adoption by a large international community, led to the development of many goal-oriented analysis methodologies [1,3,5,7,8,9]. These methodologies differ in their targeted notation and in their purpose. However, it is worth noting that most of these methodologies focus on the qualitative or/and quantitative evaluation of satisfaction levels of the goals and actors composing the model given some initial satisfaction levels [3,5,7,8,9]. Based on the *i** framework, Horkoff et al. [9] have developed

Ø. Haugen, R. Reed, and R. Gotzhein (Eds.): SAM 2012, LNCS 7744, pp. 212–228, 2013.

an interactive (semiautomated), forward propagation algorithm with qualitative values. A more recent work by Horkoff et al. [7] proposes an interactive backward propagation algorithm with quantitative values. Amyot et al. [8] have proposed three algorithms (qualitative, quantitative, and hybrid) to evaluate satisfaction levels of the intentional elements of a GRL model. Initial satisfaction levels for some of the intentional elements are provided in a strategy and then propagated, using a forward propagation mechanism, to the other intentional elements of the model through the various graph links. Giorgini et al. [5] have used an axiomatization approach to formalize goal models in TROPOS using four qualitative contribution levels (-, - -, +, ++). The authors have provided forward and backward propagation algorithms to detect three types of conflicts (weak, medium and strong). Ayala et al. [10] have presented a comparative study of *i** [2], TROPOS [5], and GRL [6]. The authors have identified (1) eight structural criteria that consider the characteristics of the language constructors, and are related to models, actors, intentional elements, decomposition elements, additional reasoning elements and external model elements, and (2) six non-structural criteria that analyze the definition of the languages, its use, and also the elements that complement them, as can be formalizations, methodologies and software tools. These criteria are syntactical.

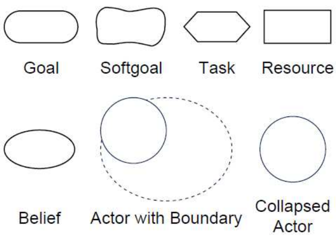
As goal models gain in complexity (e.g., large systems involving many stakeholders), they become difficult to analyze and to validate. Indeed, tentative requirements provided by the stakeholders of complex systems may be, among others, ambiguous, contradictory, and vague, which may cause many issues when the requirements engineer transforms such requirements (expressed usually in natural language) into a formal syntax in a specific goal description language. As incorrect system requirements generated from goals can lead to cost, delays, and quality issues during system development, it is essential to ensure the validity of the source goal models. Jureta et al. [1] have proposed a question-based Goal Argumentation Method (GAM) to help clarify and detect any deficient argumentation within goal models. However, their approach considers neither survey administration nor statistical analysis. To the best of our knowledge, no empirical approach has been proposed to validate goal models. In this paper, we present an approach to tackle the issue of validating complex goal models using empirical data that can be analyzed using proven statistical methods such as *Cluster Analysis* and *ANOVA (Analysis of Variance)*. We have chosen GRL [6] as target language, given its status as an international standard, but our proposed methodology can likely be applied to other goal-oriented language that visually supports actors, intentional elements, and their relationships (including i* and TROPOS), thus maintaining the discussion generic.

The remainder of this paper is organized as follows. The GRL [6] features are briefly overviewed in Sect. 2. In Sect. 3, we present and discuss the proposed GRL validation approach. Section 4 discusses how to design the validation survey from goal models. Next, empirical data analysis is presented in Sect. 5 and applied to a case study in Sect. 6. Finally, conclusions and future work are presented in Sect. 7.

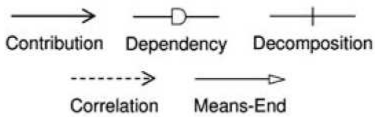
## 2 GRL in a Nutshell

The Goal-oriented Requirement Language (GRL) [6] is a visual modeling notation that is used to model intentions, business goals and non-functional requirements (NFR). GRL integrates the core concepts of:

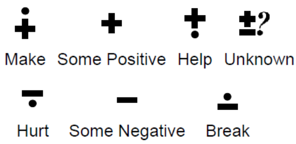
1. The NFR Framework [3], which focuses on the modeling of NFRs and the various types of relationships between them (e.g., AND, OR decomposition, positive and negative contributions, etc.). NFR comes with goal decomposition strategies along with propagation algorithms to estimate the satisfaction of higher-level goals given the attainment or non-attainment of lower-level ones.
2. The $i*$ goal modeling language [2], which has as primary concern the modeling of intentions and strategic dependencies between actors. Dependencies between actors concern goals, softgoals, resources and tasks.




(a) GRL Elements




(b) GRL Links




(c) GRL Contribution Types

**Fig. 1.** Basic Elements of GRL [8]

The basic notational elements of GRL are summarized in Fig. 1. Figure 1(a) illustrates the GRL intentional elements (i.e., goal, task, softgoal, resource and belief) that optionally reside within an actor. Actors are holders of intentions; they are the active entities in the system or its environment who want goals to be achieved, tasks to be performed, resources to be available, and softgoals to be satisfied [6]. Figure 1(b) illustrates the various kinds of links in a goal model. Decomposition links allow an element to be decomposed into sub-elements (using AND, OR, or XOR). Contribution links indicate desired impacts of one element on another element. A contribution link has a qualitative contribution type (see Fig. 1(c)) and an optional quantitative contribution. Correlation links describe side effects rather than desired impacts. Finally, dependency links model relationships between actors. For a detailed description of GRL language, the reader is invited to consult [6].

## 3 GRL Statistical Validation Approach

Figure 2 illustrates the steps of our GRL-based goal model validation approach. It is an iterative process that starts with the construction of a GRL model. In this step, the requirement engineer plays a central role in shaping the problem and solution knowledge, provided by the system stakeholders, into a GRL model. Difficulties arise when many stakeholders with different backgrounds participate in the engineering of requirements over a long period of time, which hinders the quality of the goal model. The GRL model will be used to design a validation survey (described in Sect. 4) that would be administrated to the system stakeholders (steps 2 and 3).

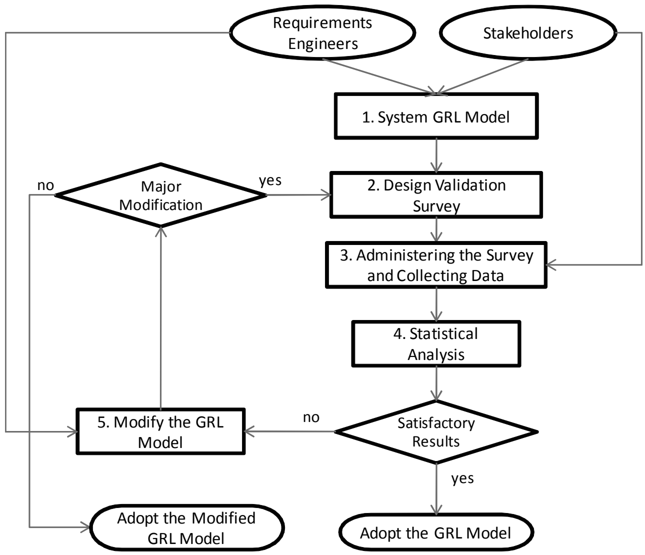



**Fig. 2.** GRL Validation Approach

Next, the resulting data is analyzed (step 4) and the identified conflicts, if any, are communicated to the involved parties. The requested modifications are incorporated into the GRL model in step 5. For major modifications, such as the deletion of many GRL elements/links or the modification of link decompositions types, we need an additional iteration. The process stops when satisfactory results are obtained.

## 4 Designing the Validation Survey

In this step, we design a survey that would be administered to the system stakeholders. Stakeholders include anyone who has an interest in the system (e.g., Customers, end users, system developers, system maintainers, etc.). The survey questions are produced based on the GRL graph intentional elements (e.g., goals, tasks, etc.), links (e.g., dependency, contribution, etc.) and constructs (e.g., AND, OR, etc.). Two types of questions may be designed:

- **Attitudinal questions** [11] typically consist of a series of statements for which stakeholders are asked to express their agreement or disagreement. A five point Likert scale [12] may be used to measure the level of agreement or disagreement. The format of a typical five-level Likert item is:
  1. Strongly agree
  2. Agree
  3. Neither agree nor disagree
  4. Disagree
  5. Strongly disagree

  The output of such questions would help the validation of the model relationships (or GRL sub-models) and would detect conflicts between stakeholders, if any.
- **Exploratory questions** are by nature open-ended as we are trying to retrieve new knowledge about a particular subject. In our strategy, we either use exploratory questions as (1) contingency questions which are administrated only in case the respondent has chosen options 3, 4, or 5 to the corresponding attitudinal question, or (2) as a simple standalone question with no prior preconditions. Exploratory questions may be completely unstructured, word association, sentence completion, story completion, etc.

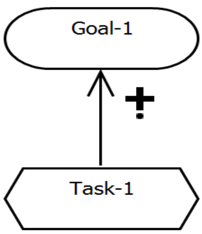
For each GRL link (e.g., dependency, contribution, etc.) or construct (e.g., AND, OR, etc.), we produce at least one attitudinal question and one or many optional contingency questions that are designed to collect pertinent information. The type and the number of contingency questions depend on the relationship that we want to validate.

Creating well structured, simply written survey questions will help in collecting valid responses. While there are no predefined rules on the wording of survey questions, there are some basic principles such as *relevance* and *accuracy* [13] that do work to improve the overall survey design. Although generic, the intent of Table 1 is to provide some tips on how to derive question vocabulary from goal model constructors. The presented examples of question vocabulary are derived from the inherent definition of goal-model constructors. However, to produce relevant, accurate, and well-understood surveys, the designed questions may include technical words from the targeted domain. Therefore, this exercise is done manually.

Figure 3 illustrates an example of a contribution relationship of type **HELP** between task *Task-1* and goal *Goal-1*, and its associated set of survey questions. In the attitudinal question, specific words are used to describe the relationship

**Table 1.** Examples of GRL Constructors and their Corresponding Questions Vocabulary

| **GRL Intentional Elements** | |
|---|---|
| **Constructor** | **Question Vocabulary** |
| Goal/Softgoal <id> | Realization/Fulfillment of Goal/Softgoal <id> |
| Task <id> | Completion/Execution of Task <id> |
| Resource <id> | Uses Resource <id> |
| Belief <text> | We believe that <text> |
| Actor <id> | Actor <id> participates |
| Actor with Boundary <id> | Actor <id> encloses |
| **GRL Intentional Relations** | |
| **Constructor** | **Question Vocabulary** |
| Make | Makes |
| Help | Helps |
| SomePositive | Has some positive contribution |
| Unknown | Has an unknown contribution |
| Hurt | Hurts |
| SomeNegative | Has some negative contribution |
| Break | Breaks |
| Decomposition AND | ... AND ... constitute |
| Decomposition OR | ... OR ... constitute |
| Dependency | Depends on |




Contribution Link of Type **HELP**

**Attitudinal Question**: "The **execution** of **Task-1** **helps** the **realization** of **Goal-1**". Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

**Contingency Question 1**: In case you don't agree with the goal **Goal-1**, please complete the sentence with an appropriate goal: "the **completion** of **Task-1 helps** the **realization** of goal ......". Otherwise rewrite **Goal-1**.
**Contingency Question 2**: In case you don't agree with task **task-1**, please complete the following sentence with an appropriate task: "The **completion** of task ...... **helps** the **realization** of goal *Goal-1*". Otherwise rewrite **Task-1**.
**Contingency Question 3**: Please complete the sentence with an appropriate verb (make, helps, has some positive contribution, hurts, breaks, etc.): "The **completion** of *Task-1* ...... the **realization** of goal *Goal-1*".

**Fig. 3.** Contribution of Type **Help** and its Associated Questions

type (i.e., verb **helps**), the involved participants with appropriate achievement description (i.e., **completion of task Task-1**, **realization of goal Goal-1**). If the respondent makes a negative answer (i.e., 3, 4 or 5), he will be asked to answer three contingency questions. In this example, we have chosen a *word completion* type of questions to check whether the issue lies within the specification of the goal (i.e., *Question 1*), within the specification of the task (*Question 2*), or within the contribution type (*Question 3*).

Deriving a question from every GRL construct/link may lead to a scalability issue when validating large models (with hundreds of links). One approach to mitigate this issue is to minimize the number of generated questions. This may be achieved by deriving questions from GRL sub-models. Section 6.1 illustrates such an optimization. For instance in decomposition links, all children are included in a sub-model. In addition, beliefs are always included in sub-models and are not assessed separately.

## 5 Validation Survey Data Analysis

Our main goal is to check whether stakeholders (survey respondents) agree on the proposed GRL model. Conflicts arise when we have major differences in the answers of the stakeholders. Therefore, such conflicts should be addressed and resolved.

Our model analysis strategy is based on the data collected from the attitudinal questions only. Contingency questions would help understand and later fix the goal model in case of negative answers to the attitudinal questions. The collected data from the attitudinal questions may be analyzed using one-way analysis of variance (one-way ANOVA), a statistical technique that can be used to evaluate whether there are differences between the mean value across several population groups. This technique can be used only for numerical data. More specifically, one-way ANOVA tests the null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

where $\mu$ = group mean and $k$ = number of groups. These variance components are then tested for statistical significance, and, if significant, we reject the null hypothesis of no differences between means and accept the alternative hypothesis (i.e., $H_1$= not $H_0$) that the means (in the population) are different from each other [14].

We use SPSS[1] software [15] to perform one-way ANOVA analysis. SPSS generates several useful tables:

– **Descriptives Table** provides useful descriptive statistics including the mean, standard deviation and 95% confidence intervals for the dependent variables for each separate group as well as when all groups are combined (see Table 2). In our context, group means relative to a question would indicate whether the groups agree (e.g., mean between 1 and 2) or disagree (e.g., mean between 4 and 5) with the corresponding GRL sub-model.

[1] Release 16.0.

- **Test of Homogeneity of Variances Table** shows the result of Levene's Test of Homogeneity of Variance (see Table 3), which tests for similar variances. If the significance value ($\sigma$, or *Sig.*) is greater than 0.05 (the $\alpha$ level of significance) then the assumption of homogeneity of variance is met and we have to look for the ANOVA Table. If the Levene's test was significant (i.e., *Sig.* is less than 0.05), then we do not have similar variances. Therefore, we need to refer to the Robust Tests of Equality of Means Table instead of the ANOVA Table.
- **ANOVA Table** shows the output of the ANOVA analysis and whether we have a statistically significant difference between our group means (see Table 4). If the significance level is less than 0.05, then there is a statistically significant difference in the group means. However, the ANOVA table does not indicate which of the specific groups differed. This information can be found in the Multiple Comparisons Table, which contains the results of post-hoc tests.
- **Robust Tests of Equality of Means Table:** Even if there was a violation of the assumption of homogeneity of variances (i.e., *Sig.* less than 0.05 in the Test of Homogeneity of Variances), we could still determine whether there were significant differences between the groups by not using the traditional ANOVA but using the *Welch* test (see Table 5). Like the ANOVA test, if the significance value is less than 0.05 then there are statistically significant differences between groups.
- **Multiple Comparisons Table** shows which groups differed from each other (*Sig.* should be less than 0.05 – see Table 6). The Tukey post-hoc test is generally the preferred test for conducting post-hoc tests on a one-way ANOVA but there are many others.

ANOVA assumes that we have a prior knowledge about the stakeholder groups and the population group to which each respondent belongs. However, when this assumption is not met, first we have to perform a cluster analysis [16], then apply ANOVA. Cluster analysis [16] is a statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. Forming groups may be inferred from additional information collected in the survey (e.g., sex, role in an organization, age, etc.). When such information is missing, the actual questions (those derived from the GRL model) may be used to form our groups.

SPSS has three different procedures that can be used to cluster data: *hierarchical cluster analysis*, *k-means cluster*, and *two-step cluster*. The *two-step* procedure is recommended for large sets of cases, whereas *hierarchical clustering* is more suitable when we want to easily examine solutions with increasing numbers of clusters. *k-means* clustering is used when we know how many clusters we want and we have a moderately sized data set.

Because we usually do not know the number of clusters that will emerge in our data set and because we want an optimum solution, a two-stage sequence of analysis may occur as follows:

1. We carry out a hierarchical cluster analysis using Ward's method applying *Squared Euclidean Distance* as the distance measure. This helps determine the optimum number of clusters we should work with. The number of clusters can be derived visually using a *dendrogram* (a hierarchical tree diagram that shows the linkage points, e.g., Fig. 12).
2. The next stage is to rerun either the hierarchical cluster analysis with our selected number of clusters, or apply k-means. This would result into allocating every case in our data set to a particular cluster.

For more details on applying cluster analysis and ANOVA using SPSS, the reader is invited to consult [16].

## 6 Illustrative Example: Introduction of a New Security Elective Course

In this section, we apply our proposed approach to a simple GRL model (see Fig. 4) that describes the introduction of a new elective course "Ethical Hacking" into the security program at King Fahd University of Petroleum & Minerals (KFUPM).

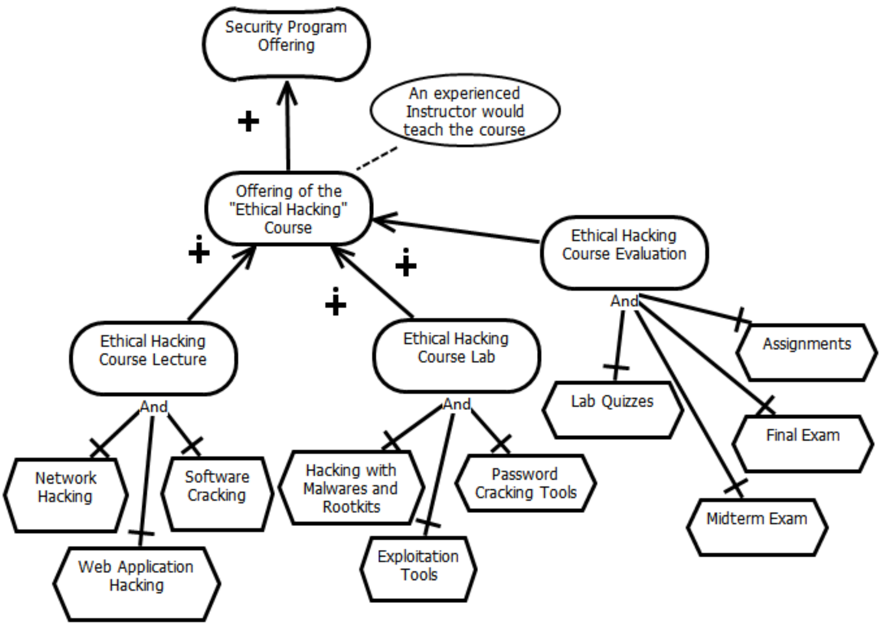
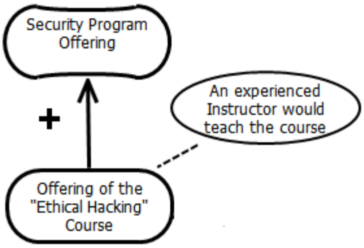



**Fig. 4.** GRL Model for the Introduction of a new Security Course

### 6.1 Designing Survey Questions

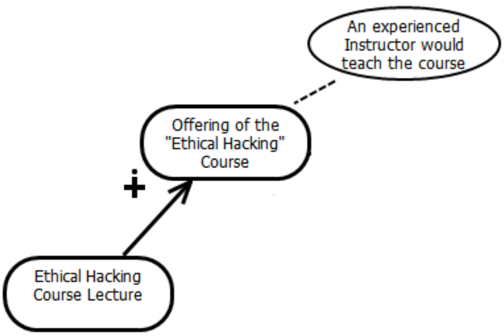
Figures 5, 6, 7, 8, 9, 10, and 11 illustrate the questions that have been derived from the GRL model artefacts. Only attitudinal questions are considered, contingency questions are out of the scope of this example. The designed survey has been administrated to the undergraduate students from the college of Computer Science and Engineering at KFUPM.




The **offering of the "Ethical Hacking" Course** has **some positive contribution** on the **realization** of the **Security Program Offering**. We believe that **An experienced Instructor would teach the course**. Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

**Fig. 5.** Question 1

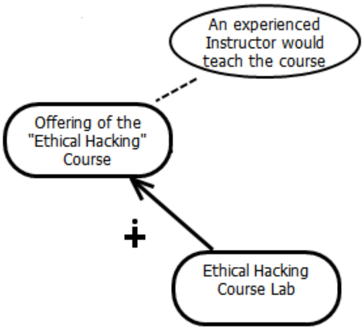



The **Ethical Hacking Course Lecture** contributes to the **offering of the "Ethical Hacking" Course**. We believe that **An experienced Instructor would teach the course**. Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

**Fig. 6.** Question 2

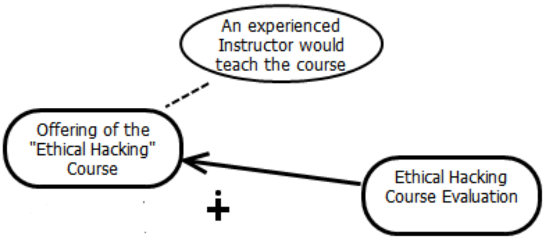
In addition to the listed questions, we have asked the students to specify their major (e.g., SWE (Software Engineering), CS (Computer Science), COE (Computer Engineering), etc.), to indicate whether they are familiar with security topics (e.g., Yes/No response), and to fill in their GPA. Since SPSS assumes that the variables are represented numerically, we have to convert the answers to (1) the major and to (2) the familiarity with security topics, to numeric values. Majors SWE, CS, and COE are converted to 1, 2, and 3 respectively while the Yes/No response is converted to 1/0.




The **Ethical Hacking Course Lab** contributes to the **offering of the "Ethical Hacking" Course**. We believe that **An experienced Instructor would teach the course**. Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

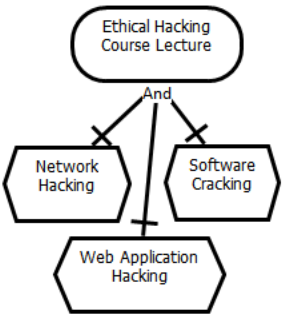
**Fig. 7.** Question 3




The **Ethical Hacking Course Evaluation** contributes to the **offering of the "Ethical Hacking" Course**. We believe that **An experienced Instructor would teach the course**. Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

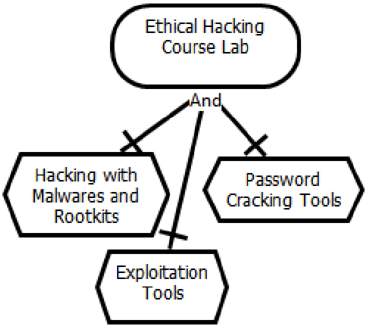
**Fig. 8.** Question 4




**Network Hacking**, **Web Application Hacking**, **<u>AND</u> Software Cracking** <u>constitute</u> the topics of the **Ethical Hacking Course Lecture**. Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

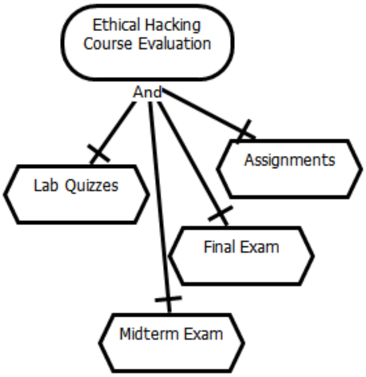
**Fig. 9.** Question 5




**Hacking with Malware and Rootkits**, **Exploitation Tools**, **AND Password Cracking Tools** constitute the topics of the **Ethical Hacking Course Lab**. Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

**Fig. 10.** Question 6




**Lab Quizzes**, **Assignments**, **Midterm Exam**, **AND Final Exam** constitute the **Ethical Hacking Course Evaluation**. Please tell us to what extent you agree with this statement?

1. Strongly agree
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Strongly disagree

**Fig. 11.** Question 7

### 6.2 Survey Data Analysis

We aim to identify conflicts between students and later clear any ambiguity about the introduction of this new course. Because we do not know the number of groups, the first step is to carry out a hierarchical cluster analysis in order to classify the 28 collected cases into distinct groups based on students' majors, GPA, and their familiarity with security topics. The application of Ward's method produces three clusters, as shown in the dendrogram in Fig. 12 (considering a reasonable linkage distance within [8,15] interval). Next, we rerun hierarchical cluster analysis with three groups which allows for the allocation of every case to a particular cluster.

Table 2 shows the group means and the standard deviation for each group. At first glance, we notice that contrary to groups 1 and 2, group 3 had a negative response to question Q7 (Mean = 4 (i.e., Disagree)).

For each question from Q1 to Q7, the Test of Homogeneity of Variances output (see Table 3) tests $H_0$: $\mu_1 = \mu_2 = \mu_3$. To interpret this output, we look at the column labeled *Sig.* This is the $p$ value. If the $p$ value is less than or equal to

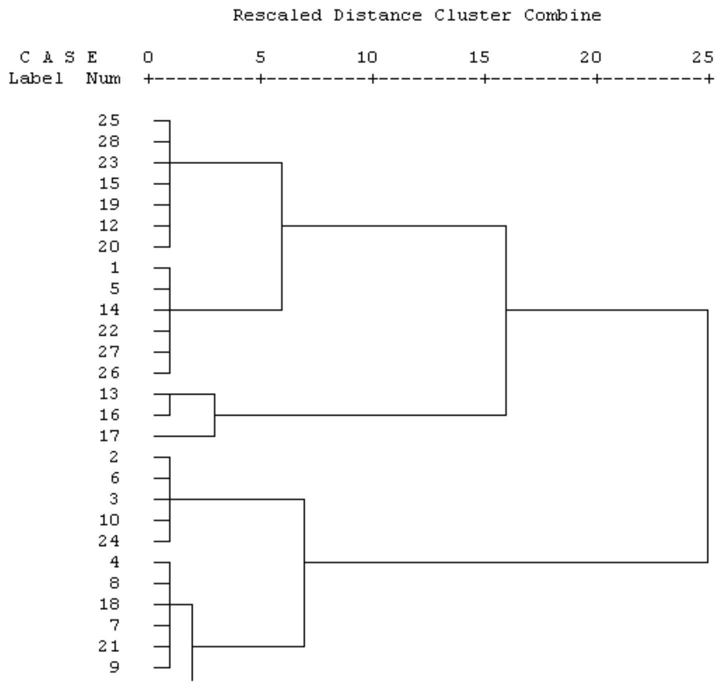



**Fig. 12.** Dendrogram Using Ward Method

the $\alpha$ level (0.05) for this test, then we can reject the null hypothesis $H_0$. If the $p$ value is greater than $\alpha$ level for this test, then we fail to reject $H_0$, which increases our confidence that the variances are equal and the homogeneity of variance assumption has been met. We can see from this case study that Levene's F Statistic have significance values of 0.613, 0.419, 0.269, 0.946, and 0.110 for questions Q2, Q4, Q5, Q6, and Q7 respectively. Therefore, the assumption of homogeneity of variance is met for these questions. Questions Q1 and Q3 (having *Sig* values of 0.024 and 0.038 respectively) do not have similar variances and we need to refer to the Robust Tests of Equality of Means Table (Table 5) instead of the ANOVA Table (Table 4).

From the ANOVA table (Table 4), we can see that the significance levels for questions Q2 (0.279), Q4(0.277), Q5(0.495) and Q6(0.822) are greater than 0.05. Hence, there is no significant differences between the groups for the underlined questions. For question Q7, the significance level is 0.00, which is below 0.05 and, therefore, there is a statistically significant difference between the three groups. The Multiple Comparisons Table (Table 6) determines which of the specific groups differed.

As stated above, there was a violation of the assumption of homogeneity of variances for questions Q1 and Q3. We could still determine whether there were significant differences between the groups by not using the traditional ANOVA

**Table 2.** Descriptive Statistics Table

| | | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound | | |
| Q1 | 1 | 13 | 1.46 | .967 | .268 | .88 | 2.05 | 1 | 4 |
| | 2 | 12 | 1.33 | .492 | .142 | 1.02 | 1.65 | 1 | 2 |
| | 3 | 3 | 2.00 | 1.732 | 1.000 | -2.30 | 6.30 | 1 | 4 |
| | Total | 28 | 1.46 | .881 | .167 | 1.12 | 1.81 | 1 | 4 |
| Q2 | 1 | 13 | 1.62 | .961 | .266 | 1.03 | 2.20 | 1 | 4 |
| | 2 | 12 | 1.67 | .985 | .284 | 1.04 | 2.29 | 1 | 4 |
| | 3 | 3 | 2.67 | 1.528 | .882 | -1.13 | 6.46 | 1 | 4 |
| | Total | 28 | 1.75 | 1.041 | .197 | 1.35 | 2.15 | 1 | 4 |
| Q3 | 1 | 13 | 1.23 | .439 | .122 | .97 | 1.50 | 1 | 2 |
| | 2 | 12 | 1.50 | .905 | .261 | .93 | 2.07 | 1 | 4 |
| | 3 | 3 | 2.67 | 1.528 | .882 | -1.13 | 6.46 | 1 | 4 |
| | Total | 28 | 1.50 | .882 | .167 | 1.16 | 1.84 | 1 | 4 |
| Q4 | 1 | 13 | 2.00 | .913 | .253 | 1.45 | 2.55 | 1 | 3 |
| | 2 | 12 | 1.67 | .888 | .256 | 1.10 | 2.23 | 1 | 4 |
| | 3 | 3 | 2.67 | 1.528 | .882 | -1.13 | 6.46 | 1 | 4 |
| | Total | 28 | 1.93 | .979 | .185 | 1.55 | 2.31 | 1 | 4 |
| Q5 | 1 | 13 | 1.85 | .801 | .222 | 1.36 | 2.33 | 1 | 3 |
| | 2 | 12 | 2.00 | 1.348 | .389 | 1.14 | 2.86 | 1 | 5 |
| | 3 | 3 | 2.67 | .577 | .333 | 1.23 | 4.10 | 2 | 3 |
| | Total | 28 | 2.00 | 1.054 | .199 | 1.59 | 2.41 | 1 | 5 |
| Q6 | 1 | 13 | 2.15 | 1.214 | .337 | 1.42 | 2.89 | 1 | 5 |
| | 2 | 12 | 2.25 | 1.288 | .372 | 1.43 | 3.07 | 1 | 5 |
| | 3 | 3 | 2.67 | 1.528 | .882 | -1.13 | 6.46 | 1 | 4 |
| | Total | 28 | 2.25 | 1.236 | .234 | 1.77 | 2.73 | 1 | 5 |
| Q7 | 1 | 13 | 1.85 | .376 | .104 | 1.62 | 2.07 | 1 | 2 |
| | 2 | 12 | 1.75 | .866 | .250 | 1.20 | 2.30 | 1 | 4 |
| | 3 | 3 | 4.00 | 1.000 | .577 | 1.52 | 6.48 | 3 | 5 |
| | Total | 28 | 2.04 | .962 | .182 | 1.66 | 2.41 | 1 | 5 |

**Table 3.** Test of Homogeneity of Variances

| | Levene Statistic | df1 | df2 | Sig. |
|---|---|---|---|---|
| Q1 | 4.322 | 2 | 25 | .024 |
| Q2 | .500 | 2 | 25 | .613 |
| Q3 | 3.754 | 2 | 25 | .038 |
| Q4 | .900 | 2 | 25 | .419 |
| Q5 | 1.383 | 2 | 25 | .269 |
| Q6 | .056 | 2 | 25 | .946 |
| Q7 | 2.413 | 2 | 25 | .110 |

but using the *Welch* test. Like the ANOVA test, if the significance value is less than 0.05 then there are statistically significant differences between groups. It is not the case since the *Welch* significance for Q1 is 0.784 and for Q3 is 0.320.

From the results so far, we know that there is significant difference between the groups for Q7 only. The *Multiple Comparisons* table (Table 6), shows which groups differed from each other. We can see that there is a significant difference between groups 1 and 3 ($P = 0.00$), and between groups 2 and 3 ($P = 0.00$). Hence, the GRL goal "Ethical Hacking Course Evaluation" *AND* decomposition needs to be reviewed involving all participants from the three groups. Once this conflict is resolved, the model might be updated, if need be.

**Table 4.** Anova Table

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Q1 | Between Groups | 1.067 | 2 | .533 | .670 | .521 |
| | Within Groups | 19.897 | 25 | .796 | | |
| | Total | 20.964 | 27 | | | |
| Q2 | Between Groups | 2.840 | 2 | 1.420 | 1.344 | .279 |
| | Within Groups | 26.410 | 25 | 1.056 | | |
| | Total | 29.250 | 27 | | | |
| Q3 | Between Groups | 5.026 | 2 | 2.513 | 3.933 | .033 |
| | Within Groups | 15.974 | 25 | .639 | | |
| | Total | 21.000 | 27 | | | |
| Q4 | Between Groups | 2.524 | 2 | 1.262 | 1.352 | .277 |
| | Within Groups | 23.333 | 25 | .933 | | |
| | Total | 25.857 | 27 | | | |
| Q5 | Between Groups | 1.641 | 2 | .821 | .723 | .495 |
| | Within Groups | 28.359 | 25 | 1.134 | | |
| | Total | 30.000 | 27 | | | |
| Q6 | Between Groups | .641 | 2 | .321 | .197 | .822 |
| | Within Groups | 40.609 | 25 | 1.624 | | |
| | Total | 41.250 | 27 | | | |
| Q7 | Between Groups | 13.022 | 2 | 6.511 | 13.630 | .000 |
| | Within Groups | 11.942 | 25 | .478 | | |
| | Total | 24.964 | 27 | | | |

**Table 5.** Robust Tests of Equality of Means

| | | Statistic[a] | df1 | df2 | Sig. |
|---|---|---|---|---|---|
| Q1 | Welch | .255 | 2 | 4.932 | .784 |
| Q2 | Welch | .582 | 2 | 5.303 | .591 |
| Q3 | Welch | 1.456 | 2 | 4.852 | .320 |
| Q4 | Welch | .758 | 2 | 5.250 | .514 |
| Q5 | Welch | 1.950 | 2 | 7.471 | .208 |
| Q6 | Welch | .134 | 2 | 5.560 | .878 |
| Q7 | Welch | 6.097 | 2 | 4.916 | .047 |

a. Asymptotically F distributed.

**Table 6.** Post Hoc Multiple Comparisons

Q7
Tukey HSD

| (I) Ward | (J) Ward | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | .096 | .277 | .936 | -.59 | .79 |
| | 3 | -2.154* | .443 | .000 | -3.26 | -1.05 |
| 2 | 1 | -.096 | .277 | .936 | -.79 | .59 |
| | 3 | -2.250* | .446 | .000 | -3.36 | -1.14 |
| 3 | 1 | 2.154* | .443 | .000 | 1.05 | 3.26 |
| | 2 | 2.250* | .446 | .000 | 1.14 | 3.36 |

*. The mean difference is significant at the 0.05 level.

## 7 Conclusion and Future Work

In this paper, we have proposed a novel GRL-based validation approach based on empirical data collection and analysis. We have applied cluster analysis and analysis of variance (ANOVA) methods in order to detect conflicts between stakeholders. Furthermore, our approach would guide argumentation and justification of modeling choices during the construction of goal models. As part of our future work, we plan to develop our survey further to go beyond conflict detection to conflict resolution.

**Acknowledgment.** The authors would like to acknowledge the support provided by the Deanship of Scientific Research at King Fahd University of Petroleum & Minerals (KFUPM) for funding this work through project No. IN111017.

## References

1. Jureta, I.J., Faulkner, S., Schobbens, P.Y.: Clear justification of modeling decisions for goal-oriented requirements engineering. Requirements Engineering 13(2), 87–115 (2008)
2. Yu, E.S.K.: Towards modeling and reasoning support for early-phase requirements engineering. In: Proceedings of the 3rd IEEE International Symposium on Requirements Engineering (RE 1997), pp. 226–235. IEEE Computer Society (1997)
3. Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J.: Non-Functional Requirements in Software Engineering. International Series in Software Engineering, vol. 5. Springer (2000)
4. van Lamsweerde, A.: Requirements engineering: from craft to discipline. In: Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering (SIGSOFT 2008/FSE-16), pp. 238–249. ACM Press (2008)
5. Giorgini, P., Mylopoulos, J., Sebastiani, R.: Goal-oriented requirements analysis and reasoning in the TROPOS methodology. Engineering Applications of Artificial Intelligence 18(2), 159–171 (2005)
6. International Telecommunication Union (ITU): Z.151 – User requirements notation (URN) - Language definition, `http://www.itu.int/rec/T-REC-Z.151/en`
7. Horkoff, J., Yu, E.S.K.: Qualitative, interactive, backward analysis of i* models. In: 3rd International i* Workshop (iStar 2008), CEUR-WS.org (2008), `http://ceur-ws.org/Vol-322/paper11.pdf`
8. Amyot, D., Ghanavati, S., Horkoff, J., Mussbacher, G., Peyton, L., Yu, E.: Evaluating goal models within the goal-oriented requirement language. International Journal of Intelligent Systems - Goal-driven Requirements Engineering 25(8), 841–877 (2008)
9. Horkoff, J., Yu, E., Liu, L.: Analyzing trust in technology strategies. In: Proceedings of the 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST 2006), Article 9. ACM Press (2006)

10. Ayala, C.P., Cares, C., Carvallo, J.P., Grau, G., Haya, M., Salazar, G., Franch, X., Mayol, E., Quer, C.: A comparative analysis of i*-based agent-oriented modeling languages. In: Proceedings of the 17th International Conference on Software Engineering and Knowledge Engineering (SEKE 2005), pp. 43–50. Knowledge Systems Institute Graduate School Calgary (2005)
11. Schuman, H., Presser, S.: Questions and answers in attitude surveys – Experiments on question form, wording, and context. Academic Press (1981)
12. Likert, R.: A technique for the measurement of attitudes. Archives of Psychology 140, 1–55 (1932)
13. Iarossi, G.: The power of survey design – a user's guide for managing surveys, interpreting results, and influencing respondents. World Bank (2006)
14. Hill, T., Lewicki, P.: STATISTICS – Methods and Applications. Statsoft Inc., `http://www.statsoft.com/textbook/`
15. IBM: SPSS software, `http://www-01.ibm.com/software/analytics/spss/`
16. Norusis, M., SPSS, I.: IBM SPSS Statistics 19 Statistical Procedures Companion. Pearson Education Canada (2011)