Representation of Latin American University Syllabuses in a Semantic Network

Mariela Tapia-Leon
Abdón Carrera Rivera
Universidad de Guayaquil
Guayaquil, Ecuador
mariela.tapial@ug.edu.ec;
abdon.carrerar@ug.edu.ec

Janneth Chicaiza Espinosa
Universidad Técnica Particular de Loja
Loja, Ecuador
jachicaiza@utpl.edu.ec

Sergio Luján-Mora
Universidad de Alicante
Alicante, España
sergio.lujan@ua.es

***Abstract* — The syllabuses describe relevant information about the teaching and learning process. Syllabuses should be created in such a way that allows automatic processing of their content to avoid typical problems of manual processing. In order to make the syllabuses intelligible to machines, firstly was necessary to create a structure based on the terms that are used in the Latin American Syllabuses. This study carried out a content analysis to extract the common terms in the syllabuses from the best universities in Latin America. Besides, we added principal aspects to make a syllabus fulfill its three core functions according to previous research. We combined automatic and manual techniques to infer the structure of a syllabus into a semantic network. The semantic network allowed describing the knowledge content of a syllabus into a graphical representation. With the representation of a semantic network, is possible to transform the network into an ontological model, whereby it is conceivable to accomplish automatic processes.**

Keywords – content analysis; semantic network; syllabus.

I. INTRODUCTION

A syllabus is a tool used by teachers and students; its main goal is to organize the way in which the teaching-learning process will be carried out during an academic course. This tool is developed by the teacher to plan an academic training route that guides, in general, how to approach such training [1].

According to Parkes and Harris [2], the three primary functions of the syllabus are: serve as a contract, serve as a permanent record, and serve as a student learning tool.

In other fields of education, the syllabus also becomes relevant. The syllabus is a document that serves as an evidence for educational control bodies to demonstrate the quality of teaching and the relevance of the topics covered in a course. Students use the syllabus to perform formalities for academic mobility. Sometimes teachers seek or request syllabuses when they want to find what other teachers have prepared for their classes to compare and improve their resources.

In this context, it is important that syllabuses are created in such a way that they allow automatic processing of their content. Thus, tasks such as the validation of subjects, the search for syllabuses, the verification of compliance of a program course and among other activities, can be done through automatic computer applications. Currently, these tasks are done manually, which sometimes leads to erroneous processing of data and unnecessary expenditure of economic and human resources.

The syllabuses describe relevant information about the teaching and learning process. Through a semantic network, it is possible to make a graphical representation of the knowledge outlined in these documents, identifying and connecting the context of the syllabus.

This study presents the creation of a semantic network model for the description of the structure of syllabuses. Two aspects have been taken into account: (1) the content currently held by the universities' syllabuses and (2) the primary functions that the syllabus must fulfill according to Parkes and Harris. Therefore, to achieve this goal we selected a subset of the syllabuses of Latin American universities that are among the 500 best in the world according to the Shanghai 2016 ranking.

From the selected corpus, we performed a content analysis to extract and organize the standard terms (categories of objects and metadata) described in the syllabuses, as a step before the creation of an ontological model. Besides, we added those terms that allow the syllabuses to fulfill the three core functions.

The rest of this paper is structured as follows, in section II, the theoretical foundations related to the syllabus and the semantic networks are presented. Moreover, the related studies are analyzed. Section III we explain the applied methodology. In section IV we present the results obtained. Finally, in section V we conclude and detail future work.

II. THEORETICAL BACKGROUND

A. Syllabus

The syllabus is a document that helps teachers to express clearly to students what they are going to learn from the course and establishes the relationship between the objectives of the course, the goals of the professional profile and the mission of the university [1].

The syllabus is important for both the student and the teacher. For the student becomes a tool that helps to guide their learning and for the teacher it behaves as a script for their classes [3].

978-1-5386-2644-3/17 $31.00 © 2017 IEEE
DOI 10.1109/INCISCOS.2017.28

According to Parkes and Harris [2], the syllabus fulfills three important functions:

- It is a contract between the teacher and students as it determines the responsibility of both parties;
- It is a registration document as it allows materializing what is taught in a course;
- It is a learning tool as it provides information that helps students build their knowledge.

Table I explains the points that a syllabus should have according to its function.

TABLE I. FEATURES OF A SYLLABUS ACCORDING PARKES AND HARRIS [2]

Syllabus as contract	Exam expiry dates Qualification weights Number and type of evaluations Course abandonment information Qualification policies Support policies Late work policies Delayed evaluation policies Warning policy on course content Class adaptation for students with disabilities
Syllabus as permanent record	Course dates Name of department Number of credits Number of hours Name and title of the teacher Prerequisites and requirements Required text information Optional text information Necessary materials for the course Objectives of the course Objectives of the course related to the professional profile Course description Description of evaluation procedures
Syllabus as a learning tool	Information to develop self-management skills Assigning time and resources to tasks Assigning time to learn subjects outside of class Tips for doing good evaluations Most common error information as reported by students Specific study strategies Time availability of teachers to attend students Teacher contact details: telephone, email, office number Information about teaching personnel or campus Information on assistance offices for students with disabilities List of websites of interest Information about the content of the syllabus and its importance for professional life

B. Semantic Network

Semantic networks arise from long-term memory studies in the field of cognitive psychology [4]. Semantic memory takes into account the human capacity to construct reality. Human interpretation allows decoding past experiences, predictions and causalities, forming new combinations of knowledge [5].

A semantic network is a graphical notation which represent knowledge in patterns of nodes and interconnected edges [6]. The nodes represent objects or concepts, and the links represent relations between the nodes [7]. The graphical representation of the nodes is formed with circles or boxes and the links with arrows or labeled edges.

The essential advantage of a semantic network is that it allows similarly representing the information as it is stored in human memory, making it also understandable by computers. This means that it is possible to analyze facts and information included in the semantic network and acquire new knowledge in an automated way [7] through automated systems.

The model of a semantic network has four components:

- Set of nodes representing real world objects.
- Set of edges representing the semantic relationships between objects.
- Set of labels that denote different types of semantic relation.
- Set of constraints that restrict semantic relationships and objects.

C. Related Studies

From the literature review, we have found some models of knowledge representation, which describe syllabuses.

Chung and Kim [8] propose an ontology to represent the learning concepts of a course and its relationships so that teachers and students can share and understand learning materials based on learning ontologies. The ontology is classified into three categories: (1) learning concepts that will be taught in the classroom; (2) Learning structure where the concepts are organized in a semantic network to describe the knowledge structures of a topic; (3) Learning materials where the teacher collects useful resources such as web pages, images, audios and videos that have a connection with the learning concepts.

The unstructured nature of syllabuses and the lack of common vocabulary between them causes search engines to produce inaccurate or irrelevant results. Yu *et al*. [9] created a digital repository of computer science syllabuses in USA. Their proposal includes modules for specific actions that help to transform an unstructured syllabus into a structured one. The named Entities Recognition Module identifies entities such as people, dates, locations, and organizations. The Topical Segmentation Module identifies the topics to be covered during course development and the qualification policy. Syllabuses are categorized according to ACM computing classification system and computer ontologies. The repository also provides a semantic search service that can assist other semantic web applications or agent-based systems.

Chung and Kim [10] designed a learning ontology that includes concepts of curriculum, achievement standards, syllabus, and learning topics. Through the proposed model it is possible to provide personalized information on a student's

learning progress based on their learning and assessment activities. The hierarchical structure of their model helps to represent what has to be studied. Students can find out their learning status and select the learning units with lower scores to review them again.

It can be concluded from the mentioned studies that in the academic community there is interest in the semantic representation of the syllabus. In summary, the relevant points from the three cited studies are: (1) the automatic processing of a syllabus is used for students and teachers to share and understand the learning materials; (2) transform an unstructured syllabus into a structured one; (3) follow up on the progress of the students. There are other interesting proposals such as those presented in [11],[12],[13].

III. METHODOLOGY

The study was carried out on a regional basis at Latin America. The Latin American universities included in the Shanghai 2016 ranking were selected. The study was not conducted with the 2017 ranking because its publication was after the elaboration of this research. However, the universities listed in Table II are still within the top 500. In the 2017 ranking the Catholic University of Chile was added however it was not considered in this work as shown in Table II.

TABLE II. SELECTED LATIN AMERICAN UNIVERSITIES

University	*Country*	*Ranking*	*Number of syllabuses*
University of Buenos Aires (UBA)	Argentina	151 - 200	3
University of Campinas (UNICAMP)	Brazil	401 - 500	3
University of Chile (UChile)	Chile	301 - 400	3
University of Sao Paulo (USP)	Brazil	101 - 150	3
Universidade Estadual Paulista (UNESP)	Brazil	301 - 400	0
Federal University of Minas Gerais (UFMG)	Brazil	301 - 400	2
Federal University of Rio de Janeiro (UFRJ)	Brazil	301 - 400	1
Federal University of Rio Grande do Sul (UFRGS)	Brazil	401 - 500	3
National Autonomous University of Mexico (UNAM)	Mexico	151 - 200	3

In order to find the syllabuses of the universities, an accurate search was made within the official website of each university using Google. The keyword that was searched was "syllabus" followed by site: www.usp.br (to cite an example). As an alternative of the word "syllabus", other terms like "curriculum", "teaching plan", "plan of the subject" with their corresponding translations in Portuguese were used. When results were not obtained, a direct search was made on the university's website. Syllabuses were also requested through communication channels such as emails and social networks. From each university, at least three syllabuses were collected because not all the faculties of the same university handle the same format and structure. Of all the universities the syllabuses were found, except in three cases (see Table II).

In this study, a content analysis was performed as a research method. "Content analysis is a particular form of document analysis. With this technique, it is not the style of the text that is intended to be analyzed, but the ideas expressed in it, is the meaning of words, themes or phrases what it tries to quantify" [14, p.173]. It was necessary to carry out an internal analysis of the documents to separate the parts of syllabuses and to know their standard elements. Then a mapping of the equivalent elements extracted from different syllabuses was conducted. The frequency of occurrence of these terms allowed to know the most common elements. According to Bardin, the calculation of data frequencies and the extraction of structures that translate into models is a controlled hermeneutics based on deduction [14].

The content analysis performed in this study followed the following steps [15]:

1. Determine the object or topic of analysis.
2. Determine the coding rules.
3. Determine the category system.
4. Check the reliability of the systems.
5. Inference.

A. Determine the Object or Topic of Analysis

The content analysis of syllabuses was carried out in order to know the common elements that the syllabuses of the best universities in Latin America have in their structure.

B. Determine the Coding Rules

The content of syllabuses from the original .doc, .pdf or HTML format was converted to .txt to determine the frequency of words. Using the NLTK library (Natural Language Toolkit) from Python, the most frequent terms were extracted, without considering stop words like articles, conjunctions, and prepositions. It was considered the three different languages of the documents: Spanish, Portuguese and English.

A code-based system was used to organize the elements of a syllabus. The Atlas.ti[1] qualitative analysis software was used to encode in each of the syllabuses. With this coding, representing the most common terms strictly related with the syllabuses' structure. Relationships were established with the goal of constructing the semantic network.

C. Determine the Category System

The categorization corresponds to determining which terms, directly and indirectly, belong to other terms forming a hierarchy. The relationship of these nodes corresponds to the form "is a".

[1] http://atlasti.com/

D. Verify the Reliability of the Systems

The reliability of the coding and the conceptual model has been validated to ensure that the data were obtained regardless who extracted the terms within the syllabus.

The authors completed a questionnaire of dichotomous questions to validate the coding system. This survey investigates the presence of terms in syllabuses (including equivalent terms or synonyms). Using the statistical tool PSPP (the free version of SPSS) the Cohen Kappa index was determined to measure the concordance of the answers of two examiners. "The procedure which suggests Cohen Kappa is that of having two (or more) judges independently categorize a sample of units and determine the degree, significance, and sampling stability of their agreement" [16, p. 37].

The whole process (B to D) has been represented in the Figure 1

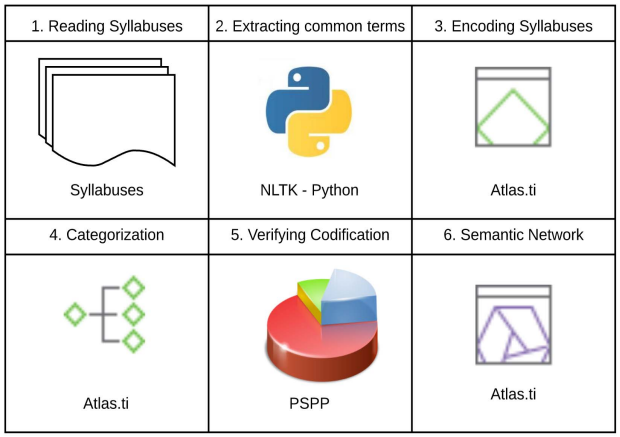

Figure 1. Process of transforming syllabuses into a unified model

E. Inference

"To infer is to explain, it is, ultimately, to deduce what is in a text" [5, p.19]. We made an inference of the syllabuses; this inference was represented in a semantic network (See Appendix A). Also, we considered aspects from Parkes and Harris' study and we added other terms to organize in a better way the semantic network (See Appendix B). The Figure 2 shows the process for making the enriched semantic network of syllabuses. This is the initial process that has been developed to create a better founded ontological model in future works.

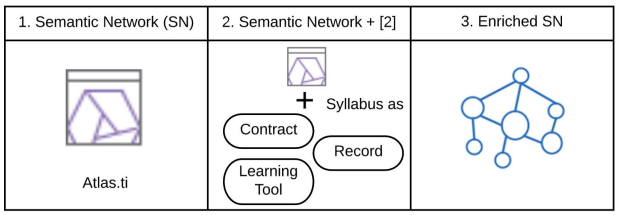

Figure 2. Process to obtain the inference

IV. RESULTS

A. About the Coding Rules

Table III shows the top-ten list of terms that appear in syllabuses, which includes words related to syllabus domain and specific topics of each document. In Table IV, we presented an extract of words that we obtained after having undergone a process of inference using Atlas.ti. With this codification was possible to create a semantic network.

TABLE III. TOP-TEN LIST OF THE MOST FREQUENT TERMS

Term	*Multilingual variations*	*Count*	*Grammatical variations*	*Count*	*Total*
Course	curso	131			178
	disciplina	41			
	course	6			
Edition	ed	98	eds	7	134
	edición	20			
	edição	9			
Student	alumno	17	alumnos	24	127
	aluno	27	alunos	42	
	estudiante	6	estudiantes	9	
	student	2			
University	universidad	36			87
	universidade	16			
	university	35			
Journal	revista	41	revistas	3	81
	journal	37			
Classroom	aula	48			67
	sala	19			
Program	programa	50	programas	4	54
Methodology	metodología	27			51
	medotologia	24			
Readings	lectura	2	lecturas	2	49
	leitura	35	leituras	3	
	readings	7			
Activity	actividad	10	actividades	35	60
	atividade	3	atividades	12	

TABLE IV. EXTRACT OF EQUIVALENT TERMS

Codes	*Frecuency*	*Equivalent Terms - Spanish*	*Equivalent Terms - Portuguese*
Course	23	Disciplina, Asignatura, Curso	Disciplina, curso
Professor	21	Profesor, profesor invitado, profesor anfitrión, equipo de profesores, docentes	Professores, profa, professora, corpo docente
Program	21	Temario, programa, estructura del seminario, contenido	Conteúdo programático, tópicos do programa, conteúdos

Codes	*Frecuency*	*Equivalent Terms - Spanish*	*Equivalent Terms - Portuguese*
Assessment	20	Evaluación, sistema de evaluación, acreditación, nota final	Avaliação, criterios de avaliação, avaliação e relatório
Objetive	18	Objetivo, objetivos	Objetivos da disciplina, objetivos do curso, escopo e objetivo
University	18	Universidad	Universidade
Bibliography	15	Lecturas, bibliografía, texto de estudio,	Referências bibliográficas, leituras, leituras prévias obrigatórias
Course description	14	Fundamentos, justificación, descripción del curso	Ementa, Descrição do curso, Súmula
Methodology	14	Procedimientos, Desenvolvimiento, Metodología de trabajo, principios metodológicos, estrategia pedagógica	Estrutura do curso, metodología de ensino
Faculty	13	Facultad	Facultade
Schedule	13	Cronograma, horario, horarios	Horários
Basic bibliography	11	Bibliografía básica, textos bases, bibliografía de base, bibliografía básica y obligatoria	Referências básicas
Course code	10	Código	

B. About Category System

The hierarchy was established based on the semantic network that was built in Atlas.ti. An example of a hierarchy is represented in the Figure 3.

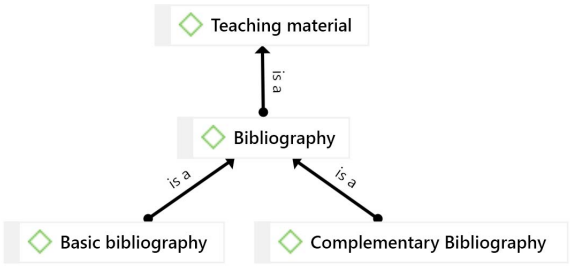

Figure 3. Example of a category system

C. About Verify the Reliability of the Systems

Two examiners reviewed the existence of terms in the syllabuses to determine the reliability of the coding system. Table V shows the Cohen Kappa index (K) obtained after processing the responses to the survey applied. The first column indicates the name of the university and the following columns show the matching of K answers for each of the three syllabuses analyzed (S1, S2, S3).

As shown in Table VI, the agreement of the answers between the two examiners is good and very good. That ratifies that the coding system is reliable.

TABLE V. STATISTICAL ANALYSIS KAPPA DE COHEN

University	*KS1*	*KS2*	*KS3*
UBA	0.93	0.85	0.78
UNICAMP	0.84	0.84	0.78
UCHILE	0.86	0.79	0.70
USP	0.86	0.79	0.66
UNESP	X	X	X
UFMG	0.93	0.92	X
UFRJ	0.78	X	X
UFRGS	0.93	1.00	0.93
UNAM	0.92	1.00	1.00

TABLE VI. INTERPRETATION OF THE COHEN KAPPA INDEX

Cohen Kappa Index	*Interpretation*
0.00 - 0.20	Lowest concordance
0.21 - 0.40	Poor concordance
0.41 - 0.60	Moderate concordance
0.61 - 0.80	Good concordance
0.81 - 1.00	Very Good concordance

D. About the Inference

The inference that we made was constructed from a subset of syllabuses and was represented by a semantic network model. However, as can be seen in Table VII, the analyzed documents are not complete according to the features proposed by Parkes and Harris [2]. To check the percentage of compliance of the syllabus features according to this approach, a comparison between the aspects enlisted in Table I and the content of the syllables was performed. The results are shown in Table VII.

TABLE VII. COMPLIANCE OF SYLLABUS FEATURES

Country	*University*	*Format file*	*Percentage of compliance*			
			Contract	*Permanent record*	*Learning tool*	*Average by university*
Brasil	UFRGS	pdf	12.1	84.8	3.0	**33.3**
Brasil	UFRJ	pdf	27.3	63.6	0.0	**30.3**
Brasil	SaoPaulo	pdf	39.4	93.9	33.3	**55.6**
Argentina	UBA	pdf	12.1	78.8	24.2	**38.4**
Chile	UCHILE	pdf	42.4	60.6	27.3	**43.4**
México	UNAM	pdf	27.3	51.5	21.2	**33.3**
Brasil	UFMG	html	9.1	90.9	18.2	**39.4**
Brasil	UNICAMP	doc	12.1	90.9	9.1	**37.4**
Average by function			**22.7**	**76.9**	**17.0**	**38.9**

Considering the above explained, authors proceeded to identify each of the elements proposed by Parkes and Harris [2] and that had not been described by any syllabus. Table

VIII shows the new terms that have been added to the syllabus.

TABLE VIII. NEW TERMS ADDED TO SEMANTIC NETWORK

New Term	*Description*
FAQ	Syllabus as a learning tool
Self-management skills	
Assessments tips	
Study strategy	
Expiry dates	Syllabus as contract
Assessment type	
Number of assessment	
Accessibility	
Qualification	
Late work	
Delay Assessment	
Drop out	
Content warning	
Support	
Description of assessment	Syllabus as permanent record

With the terms extracted from the syllabuses, the terms added from the main functions according to Parks and Harris, and other words added to organize, we have made an enriched semantic network of syllabuses. The new terms are represented in gray boxes like is showed in Appendix B.

V. CONCLUSIONS

In this study, the authors have followed a systematic process based on content analysis. A mix of automatic and manual techniques assisted them to obtain the common terms into the syllabus of best universities in Latin America.

NLTK library allowed to obtain the frequency of the words. This automatic process counted unnecessary words like "edition" in the bibliographic section. Thus, was necessary an inference process to determine which terms were strictly related to the structure rather than the content. Atlas.ti allowed making that distinction.

As can be seen in Table VII, the most common use of the syllabuses is like a record of basic information of each course as bibliography, objectives and professor's name. A syllabus from contract dimension is less usual; in this case, information about the number of exams and the kind of assessments are the most filled. Meanwhile, policies related to delayed assessments, courses' abandonment or interruption and content accessibility never are described. Finally, a syllabus seen as a learning tool is the dimension less explored. Only information related to the assessment such as percentage, score or the number of exams is declared into a syllabus. With this argument, the authors have made an enriched semantic network based not only in the common terms extracted of the syllabuses, besides, they have also incorporated terms that are considerate important to look a syllabus as a contract, a permanent record, and a learning tool.

All this process was made with the purpose to have a semantic network of the syllabuses that will be represented in a machine language like .owl in future works. Thus it will be possible automate some processes with the syllabus that currently carries out manually.

REFERENCES

[1] N. C. De Camacho, "El Syllabus," *El Tecnológico*, vol. 23, no. 1, p. 22, 2013.

[2] J. Parks and M. Harris, "The purpose of a syllabus," College Teaching, vol. 50, no. 2, pp. 55–61, 2002.

[3] J. Chicaiza, N. Piedra, and J. López-Vargas, "Diseño de un vocabulario para conectar e interoperar los syllabus de cursos OCW mediante tecnologías de la Web Semántica," in *Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2014)*, 2014.

[4] E. Tulving, "Episodic and semantic memory," *Organ. Mem.*, vol. 1, pp. 381–403, 1972.

[5] J. Á. Vera-Noriega, C. E. Pimentel, and F. J. Batista de Albuquerque, "Redes semánticas: aspectos teóricos, técnicos, metodológicos y analíticos," Ra Ximhai, vol. 1, no. 2003, pp. 439–451, 2005.

[6] J. F. Sowa, "Semantic Networks," in *Encyclopedia of Cognitive Science*, Chichester: John Wiley & Sons, Ltd, 2006.

[7] A. Fajar Santoso, I. Supriana, and K. Surendro, "Designing Knowledge of The PPC with Semantic Network," *Journal of Physics: Conference Series*, vol. 801, no. 1, p. 12015, Jan. 2017.

[8] H.-S. Chung and J.-M. Kim, "Semantic Model of Syllabus and Learning Ontology for Intelligent Learning System," in *International Conference on Computational Collective Intelligence*, 2014, pp. 175–183.

[9] X. Yu *et al.*, "Using Automatic Metadata Extraction to Build a Structured Syllabus Repository," in *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 337–346.

[10] H. Chung and J. Kim, "A New Personalized Syllabus Model based on Achievement Standards Analysis and Evaluation," in *Proceedings of the World Congress on Engineering and Computer Science 2016*, 2016.

[11] H. Chung, S. Lee, and J. Kim, "Learning Concept Sequencing through Semantic-based Syllabus Design and Integration," *Indian Journal of Science and Technology*, vol. 8, no. 18, Aug. 2015.

[12] A. Petiwala and K. Moudgalya, "Constructing a learner centric semantic open syllabus for automated text book generation," in *2012 IEEE International Conference on Technology Enhanced Education (ICTEE)*, 2012, pp. 1–9.

[13] R. Ramesh, S. M., and S. Iyer, "Annotating the Domain Ontology of a Course with ITS Syllabus and Learning Objectives," in *2016 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*, 2016, pp. 130–131.

[14] F. Lopez Noguero, "El análisis de contenido como método de investigación," *Revista Internacional de Investigación e Innovación Pedagógica*, vol. 4, pp. 167–179, 2002.

[15] J. Andréu Abela, "Las técnicas de Análisis de Contenido: Una revisión actualizada.," *Fundación Centro Estudios Andaluces, Universidad de Granada*, vol.10, no. 2, pp. 1-34, 2000.

[16] J. Cohen, "A Coeficient of Agreement for Nominal Scales," *Educational and Psychological and Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

APPENDIX A

APPENDIX B