

リレーショナルデータモデル演習システムのための 正規化問題生成機能の開発と運用評価

岡田信一郎^{†1,a)} 阿部孝昭^{†2} 山縣大輔^{†3}

概要：現在、情報分野を専門とする大学生にとってリレーショナルデータベースの基礎理論となるリレーショナルデータモデルの習得は必須といえる。しかし、この理論、とくに正規化を苦手とする学生は多い。そこで筆者らは演習を数多く行うことでリレーショナルデータモデルを習得するための「リレーショナルデータモデル演習システム」を開発し、運用評価を行ってきた。このシステムはリレーショナル代数および正規化に関する問題を自動生成し、学習者解の正誤を判定する機能を備えており、学習者は多数の問題による演習を行うことができる。今回はこのシステムの開発・改良と3年間の運用評価の結果、とくに正規化問題生成機能に対する(1)関数従属性の厳密化と(2)正規形を問う問題の追加の2点の改良とその結果について報告する。

キーワード：リレーショナルデータモデル、演習支援、正規化、問題生成

1. はじめに

現在、情報分野を専門とする技術者にとってリレーショナルデータベースの基礎理論となるリレーショナルデータモデルの習得は必須といえる。しかし、データベース技術を学習する学習者にリレーショナルデータモデル、とくに正規化理論を苦手とする者は多い。例えば、本学情報工学科の2012年度「データベース論」の定期試験において、SQLの問題に対する平均得点率が86%であったのに対し、リレーショナルモデルに関する問題の平均得点率は57%、正規化問題に限れば41%であった。

このような学習者に対しては簡単な問題を多数解かせることによって、理論の理解を促すことが考えられる。しかし、多数の問題を作成し、学習者の解答の正誤を判定することは、指導者にとって大きな負担となる。

そこで筆者らは、リレーショナルデータモデルの理解向上を目的とした、リレーショナルデータモデル演習システムの開発を進めている。

このシステムは、WWWブラウザから利用することができ、リレーショナル代数8種の問題と、関数従属性に基づく正規形である第2正規形からボイスコード正規形までの3種の正規化問題の自動生成、正誤判定を行うことができる。

リレーショナルデータモデルの学習を支援するツールとしては、sAccess [1]などが提案されている。sAccessはデータベース初学者、とくに情報を専門としない学習者を対象としており、データベースのテーブルへの操作を画面上で確認しながら対話的に行えることが特徴といえる。

これに対し筆者らは、情報を専門とする学習者にリレーショナルデータモデルの理論を正確に習得させることを目指したため、リレーションに対する操作結果を学習者に正確に解答させるシステムを開発した。またシステム開発の技術的な課題として不自然さのない問題リレーションを自動生成する手法の確立を挙げている。

筆者らのシステムは2013年度までに基本機能の実装を終え、実際の運用を通じた機能の検証を行っている [2]。その結果、出題数、難易度ともに適切で、成績の向上に寄与していることが確認されたが、正規化問題で自動生成され提示されるリレーションの不自然さが問題として残った。

そこで、2014年度には正規化問題で出題されるリレーション生成のための仕組みを見直し、正規形の条件となる関数従属性をひとつおりの検証したテンプレートを使用する方法へ変更した。 [3]。その結果、問題リレーションの行数が増えたものの、問題の難易度には大きな影響がなく、不自然さを解消することになった。

しかし、その後の見直しにより、テンプレートの一部に関数従属性検証の漏れがあったため、2015年度はさらに修正したテンプレートにより運用を行っている。また、2015

^{†1} 現在、茨城大学

^{†2} 現在、株式会社システムズ

^{†3} 現在、北茨城市

a) shinichirou.okada.mzfe2@vc.ibaraki.ac.jp

年度は学習者が正規形の条件を正確に理解できるようにするために、正規化問題に正規形を問う問題を追加することを試みている。

本稿では、リレーショナルデータモデル演習システムの正規化問題生成機能について、上述の改良を中心に、3年間の運用を通じた評価の結果を報告する。

2. リレーショナルデータモデル演習システムの機能概要

筆者らが開発、運用を行っているリレーショナルデータモデル演習システムは以下の機能によって構成される。

- 問題生成機能
 問題を要求された際に問題文、問題リレーション、解答リレーションを生成する機能。
- 問題表示機能
 WWW ブラウザに問題を表示する機能。
- 解答入力機能
 学習者に解答を入力させる機能。
- 正誤判定機能
 学習者の解答と解答例リレーションを比較し、それらが同一であることを判定する機能。
 正誤判定の流れを以下に示す。
 1. 解答例と学習者解の行数、列数を比較。
 2. 学習者解の列を、解答例の並びと同じになるように変形し、列名の並びを比較。
 3. 解答例と学習者解を行単位で全て比較し、一致しない行が存在しなければ正解となる。
- 正解表示機能
 正誤判定からの結果と解答例を表示する正解表示機能。
- 学習履歴管理機能
 学習者が学習した問題の履歴を管理する履歴表示機能。詳細を表示すると、問題文、問題リレーション、解答リレーション、解答例リレーションが表示される。
- 成績管理機能
 学習者の成績を管理する成績管理機能。
 1つの学習項目につき3問以上の正解、または6問以上問いた場合、最近5問の正答率が60%(3問)以上正解で合格とする。

学習項目は、「リレーショナル代数」から和、差、直積、共通、商、射影、選択、結合の8種類の演算、「正規化」から第2、第3、ボイスコード正規形の3種類の正規化問題で、合わせて11項目となる。学習者は、「リレーショナル代数」か「正規化」までを選択できるが、あとはシステムによってランダムに学習項目が選択、出題される。

図??に、リレーショナルデータモデル演習システムの画面表示の例を示す。

本システムは、2013年度から2015年度の3年間の運用の間に改良が続けられている。以下に、本システムの各年

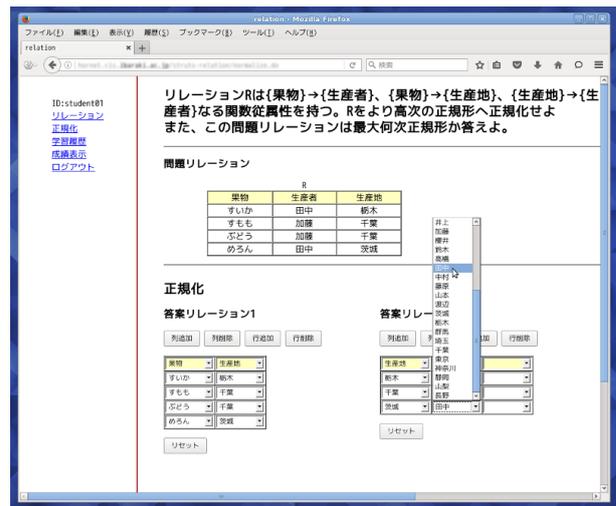


図 1 リレーショナルデータモデル演習システムの画面表示例

表 1 関数従属性列 x 列 y が存在するリレーション R

R		
x	y	z
x_0	y_0	z_0
x_0	y_0	z_1
x_1	y_0	z_2
x_2	y_1	z_3
x_2	y_1	z_4

度における特徴や改良について説明する。

2.1 2013年度の実装の特徴

2013年度の正規化問題生成は、問題リレーションに必要な関数従属性を満たす値のみを適切に設定し、その他の値はランダムに設定する方法で行っていた。例えば、第2正規形への正規化問題の場合には、第2正規形違反となる部分関数従属となる列^{*1}を決め、それらの列には部分関数従属となるような値を設定し、その他の列にはランダムに値を設定し、 4×4 程度の大きさのリレーションにしている。

なお、リレーション R の列 x から列 y への関数従属性 ($\{x\} \twoheadrightarrow \{y\}$) とは、列 x のある値に対して列 y の値が1種類しか存在しない関係が常に成り立つことをいう [4]。例えば、表1に示すリレーション R の列 x から列 y に関数従属性があるとする。このリレーション R の列 x の値が x_0 のときには、列 y の値は必ず y_0 となっている。一方、列 x が x_0 のとき、列 z の値は z_0 と z_1 の2種類があり、列 x から列 z には関数従属がないことが明らかである。

なお、関数従属性がない列間にはどのような値の組み合わせが現れてもよい。表1のリレーション R の列 z には同じ値の行が存在していないため、列 z から列 x に関数従属性が存在しないことは明らかではないが、関数従属性が

^{*1} リレーショナルデータモデルの理論では「ダブル」、「属性」の用語が用いられるが、本稿では説明のしやすさから、それぞれに「行」、「列」の用語を使用する。

存在する条件も満たしてはいない。そのため、問題リレーションの生成においては、関数従属している列間の値の組み合わせが正しければ、残りの列はランダムに値を決めても問題としては成立する。

図 2 に 2013 年度のシステムによって生成された正規化問題の例を示す。この例では、関数従属性 $\{w\}$ $\{z\}$ があるように見えるが、問題文に記述された関数従属性とは無関係であり、関数従属ではない。

リレーション R は $\{w,x\}$ $\{y\}$ 、 $\{w,x\}$ $\{z\}$ 、 $\{x\}$ $\{y\}$ なる関数従属性を持つ。R をより高次の正規形へ正規化せよ。

R

w	x	y	z
a	i	m	s
a	k	p	s
b	j	o	r
e	k	p	r
f	k	p	r

図 2 2013 年度システムによって生成された問題の例

しかし、運用後のアンケートにおいて、リレーション内の関数従属性の不自然さを問題点に挙げる学習者が 9 名 (全体の 20%) 存在し、改善点とされた。

また、2013 年度の実装では、問題リレーション中の列名、値はすべて 1 文字の英数字、またはギリシャ文字となっており、それらをキーボードから入力しなければならないユーザーインターフェースの使いにくさが指摘されている。

2.2 2014 年度の改良

2014 年度は、前節の正規化問題の問題リレーションに存在しない関数従属性が存在するように見える問題点を解消するために、あらかじめ関数従属性の検証を行ったテンプレートにランダムに値を当てはめる方法で問題リレーションを生成する方法に改めた。

前節の表 1 のリレーション R は、列 x が x_0 のとき、列 z の値は z_0 と z_1 の 2 種類があり、列 x から列 z には関数従属性がないことが明らかであった。関数従属性がない列間にこのような値の組み合わせが必ず存在するようにすれば、問題文で指定しない関数従属性が問題リレーションに存在しないことを明確にできる。

そこで、以下の手順で問題リレーション生成のテンプレートを生成した。

- (1) すべての列の組み合わせを列挙し、関数従属性の有無をチェック。
- (2) 関数従属性の有無に応じた値のパターンを設定。
- (3) そのままでは数十行のリレーションになってしまうので、重複する値のパターンをまとめて行数を削減。

この手順で、2014 年度版システムのために 7 種類のテン

プレートを作成した。その一覧を表 2 に示す。2014 年度と 2015 年度の行数の違いは、次節で説明する。それぞれの正規形への正規化問題リレーションの生成は、ランダムに選択したテンプレートへ値を当てはめることで行われる。

表 2 正規化問題リレーション生成用テンプレートの種類と行数

学習項目	列数		行数	
	全体	主キー	2014 年度	2015 年度
第 2 正規化	3 列	2 列	6 行	10 行
第 2 正規化	4 列	2 列	8 行	
第 3 正規化	3 列	1 列	4 行	12 行
第 3 正規化	4 列	1 列	8 行	
第 3 正規化	4 列	2 列	10 行	
ボイスコード正規化	3 列	2 列	6 行	
ボイスコード正規化	4 列	2 列	10 行	

テンプレートの例を表 3 に示す。このテンプレートは、3 列のリレーションを第 2 正規形へ正規化する問題のテンプレートであり、関数従属性 $\{x, y\}$ $\{z\}$ 、 $\{x\}$ $\{z\}$ がある。この 2 つの関数従属性に由来する関数従属性および自明な関数従属性以外の関数従属性が存在しない値の組み合わせから作られている。

表 3 第 2 正規化問題の 3 列のテンプレート

x	y	z
x_0	y_0	z_0
x_0	y_1	z_0
x_1	y_2	z_1
x_2	y_2	z_1
x_3	y_3	z_2
x_4	y_3	z_3

2014 年度はこれらのテンプレートを用いた正規化問題生成機能をシステムに実装し、運用を行った。

また、2014 年度は、英数字、ギリシャ文字 1 文字だけからなる問題リレーションの要素に「りんご」「茨城」のような一般的な単語も使えるようにする改良も行った。さらに解答の入力をキーボード入力から、プルダウンメニューからの選択式に変更した。図 1 に示した画面表示例は、これらの改良以降のものである。

2.3 2015 年度の改良

2014 年度の運用評価後の確認の結果、2 つのテンプレートに関数従属性がない値の組み合わせの検証漏れがあることが確認された。2015 年度はこの誤りを修正し、運用を行った。

表 2 で 2015 年度に行数が変化しているものが修正されたテンプレートである(「」は変化なし)。表 4 に、第 3 正規化 4 列 (主キー 2 列) のテンプレートの 2014 年度版と 2015 年度版の両方を示す。この修正により、問題リレーションの最大行数が 12 行と、かなりの大きさになった。

表 4 第 3 正規化 4 列 (主キー 2 列) テンプレートの修正

2014 年度				2015 年度			
w	x	y	z	w	x	y	z
w0	x0	y0	z0	w0	x0	y0	z0
w0	x1	y0	z0	w0	x1	y0	z0
w1	x2	y1	z1	w1	x2	y1	z1
w1	x3	y2	z2	w1	x3	y2	z1
w2	x4	y3	z3	w2	x4	y3	z2
w3	x4	y4	z4	w2	x5	y4	z3
w4	x5	y5	z5	w3	x6	y5	z4
w5	x5	y6	z5	w4	x6	y5	z4
w6	x6	y7	z6	w5	x7	y6	z5
w7	x6	y7	z6	w6	x7	y7	z5
				w7	x8	y8	z6
				w8	x8	y9	z7

システム導入初年度の 2013 年度を含めた問題リレーションの行数、列数の変化を表 5 に示す。関数従属性を厳密に検証したテンプレートを使用した結果、問題リレーションの行数が大きく増加した。

表 5 正規化問題リレーションの行数、列数の変化

学習項目	2013 年度	2014 年度	2015 年度
第 2 正規化	(4,3) ~ (5,4)	(6,3) ~ (8,4)	(6,3) ~ (10,4)
第 3 正規化	(4,3) ~ (5,4)	(4,3) ~ (10,4)	(4,3) ~ (12,4)
ボイスコード正規化	(4,3) ~ (5,4)	(6,3) ~ (10,4)	(6,3) ~ (10,4)

(m, n) は m 行 n 列を表す。

また、これまでのシステムの正規化問題は提示された問題リレーションがどのような正規形になっているかを問わなかった。そのため、各正規形の条件を正確に覚えていなくても、主キー以外からの関数従属性に注目すれば、正規化問題を解くことが可能であった。このことを解消するために、2015 年度は問題リレーションがそのような正規形であるかを問う機能を実装し、正規化問題の一部として追加した。

この改良によって正規化と正規形の両方の解答が正しくなければ不正解となるため、難易度は高くなった。

図 3 に正規形を問う問題の表示例を示す。

3. 運用による評価

3.1 運用評価の概要

2013 年度から 2015 年度に渡って、本学情報工学科 2 年生向けに開講されている「データベース論」の受講者に、授業の課題として本システムを使用させた。リレーショナルデータモデルに関する説明は授業中に行い、演習期間内の授業中の演習時間または空き時間に演習を行うように指示している。また、演習期間の終了後に、アンケートを実施している。表 6 に各年度の演習期間、学習履歴の研究協力

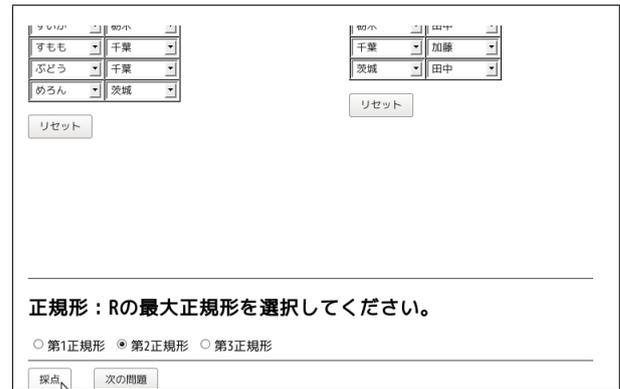


図 3 正規形を問う問題の表示例

に同意した学習者の人数、アンケートへの回答数を示す。

表 6 各年度の演習期間と学習者数

年度	期間	学習者数	アンケート回答数
2013	11 月 19 日から 12 月 11 日	51	44
2014	11 月 18 日から 12 月 12 日	47	35
2015	11 月 24 日から 12 月 15 日	72	47

学習履歴から得られたシステム全体 (リレーショナル代数も含む) のデータを表 7 に示す。また、表 8、9 に問題数、問題の難易度に関するアンケート結果を示す。

各年度とも、1 人あたりの解答数 (出題数) が 40 問程度、正答率が 70% 程度で、アンケートに対しても、問題数が「ちょうど良い」、難易度が「普通、ちょうどいい」が最も、出題数、難易度は適切であったと考えられる。

表 8 アンケート設問「問題数は適切でしたか？」への回答

選択肢	回答数		
	2013 年度	2014 年度	2015 年度
少なかった	4	2	1
多かった	2	5	6
ちょうど良かった	38	28	40

表 9 アンケート設問「リレーショナル代数、正規化問題の難易度はどうでしたか？」への回答結果

選択肢	回答数		
	2013 年度	2014 年度	2015 年度
易しい	2	3	0
普通、ちょうどいい	34	27	37
難しい	8	5	10

3.2 正規化問題の不自然さの評価

表 10 に正規化問題の関数従属性の不自然さに関するアンケートの回答を示す。なお、設問の文章は、2013 年度は「現状のシステムでは正規化の問題において関数従属性があると言われていないところにあるように見えてしまう場合があります。この場合、問題を解くにあたって気にな

表 7 システム全体の学習履歴データ

項目	2013 年度	2014 年度	2015 年度
学習者数 (人)	51	44	71
総解答数 (件)	2128	1932	3134
総正解数 (件)	1619	1443	2327
正解率 (%)	76.1	74.7	74.3
1 人当たりの解答数 (件)	41.7	43.9	44.1

りましたか?」2014 年度以降は「正規化問題において、リレーション中の値は、問題文に提示されていない関数従属性が存在しないように考慮されています。不自然さを感じたところはありませんか?」と問題生成法の違いにより異なっている。

表 10 正規化問題の関数従属性の不自然さに関するアンケートへの回答

選択肢	回答数		
	2013 年度	2014 年度	2015 年度
気になった/あった	9	0	1
気にならなかった/なかった	34	28	44
未回答	1	7	2

システム導入初年度の 2013 年度には関数従属性の不自然さが気になる学習者が 9 人いた。これは全体の 20% に及ぶ。そこで、2.2 節で説明した改良を行った結果、2014 年度以降、関数従属性の不自然さを指摘する学習者は少なくなった。しかし、一方で 2.2 節の表 2、2.3 節の表 5 に示したとおり、問題リレーションの行数が増える結果となった。

関連するアンケートの結果を表 11 と表 12 に示す。表 11 はシステム全体で問題リレーションの行数や列数が適切であったかどうかを問う設問の回答である。なお、2015 年度は選択肢の表現が変更されているので、それ以前の選択肢との併記となっている。表 11 によれば、システム全体としては、年々問題リレーションの行数、列数が増えていると感じている学習者が増えている。同じ設問で、具体的にどの学習項目の行数や列数が多いと感じているのか回答した結果を示すのが表 12 である。この表によると、正規化問題の行数や列数が多いと解答する学習者の数は、2013 年度は 0 であったが、2014 年度には急激に増えている。正規化問題の問題リレーションの行数が増えたことの影響と考えられる。さらに 2015 年度には正規化問題のリレーションの行数や列数を多いとする学習者の数は増えているが、2015 年度はアンケート回答数も、行数が多いと回答した人数も多く、それらとの比においては 2014 年度から多く増加しているとは言えず、テンプレート修正による影響は確認できなかった。

3.3 正規形を問う問題の追加の評価

表 13 に、2014 年度と 2015 年度の学習履歴から得られ

表 11 アンケート設問「問題リレーションの行数や列数は適切でしたか?」への回答

選択肢	回答数		
	2013 年度	2014 年度	2015 年度
少なかった/少ないと思うトピックがある	1	0	0
ちょうど良かった/普通、ちょうどいい	40	24	21
多かった/多いと思うトピックがある	3	11	25
少ない、多いと思うトピックがそれぞれある	—	—	1

表 12 行数や列数が多いと感じられた学習項目

選択肢	回答数		
	2013 年度	2014 年度	2015 年度
和演算	0	2	1
差演算	0	2	0
直積演算	1	2	6
共通演算	0	0	2
射影演算	0	2	0
選択演算	0	2	0
結合演算	2	2	2
商演算	0	2	0
第 2 正規化	0	10	16
第 3 正規化	0	10	16
ボイスコード正規化	0	10	16

た正規化問題の解答数と正答率を示す。2014 年度は正規形を問う問題はないので、問題そのものの正答率と正規化の正答率は同じである。

2015 年度に難易度は高くなっているが、正答率の低下は少ない。これは各学習項目が正答率 60% を合格基準としているため、60% に近い値が維持されたと考えられる。さらに学習者 1 人あたりの解答数にもほとんど変化がなかった。ただし、2015 年度の正規化のみ、正規形のみ、正答率をみると、10 ポイントほど高くなっている。正規化と正規形の両方を回答することによって問題の難易度は高くなったが、学習の密度が高くなったことで同じ問題数で同程度の正答率に達したものと考えられる。

3.4 その他の改良の評価

2014 年度には「入力を選択式にする」「正規化問題のリレーションの要素に一般的な単語を加える」改良を行っている。表 14、15 にそれらの改良に対するアンケートの結果を示す。なお表 14 の設問は「正規化問題において、今回要

表 13 正規化問題の解答状況

項目	解答数・正答率					
	2014 年度			2015 年度		
	第 2 正規形	第 3 正規形	ボイスコード	第 2 正規形	第 3 正規形	ボイスコード
解答数	220	207	232	349	346	368
1 人あたりの解答数	5.0	4.7	5.3	4.9	4.9	5.2
正答率 (%)	65	69	61	59	61	55
正規化のみの正答率 (%)	65	69	61	70	71	61
正規形のみの正答率 (%)	—	—	—	80	83	80

素がアルファベットなどの記号である問題と、りんごや茨城などの具体的な値である問題があったと思いますが、どちらの方が解答しやすかったですか?」表 15 の設問は「今回から解答入力部分を選択方式にすることで入力ミス軽減を目指しましたが、解答しやすかったですか?」である。

表 14 リレーションの要素に関するアンケート回答

選択肢	回答数	
	2014 年度	2015 年度
アルファベットなどの記号	9	11
具体的な値	16	17
どちらでも良い	3	17
未回答	7	2

表 15 選択式入力に対するアンケート回答

選択肢	回答数
解答しやすかった	23
解答しにくかった	5
未回答	7

表 14 によれば、リレーションの要素は具体的な値（一般的な単語）の方が良いとする学習者が多いが、アルファベットなどの記号の方が良いとする学習者も少なくはない。本システムの現状では、どちらかに統一することは好ましくないと考えられる。

表 15 からは、解答リレーションの入力を選択式した結果、解答の入力が容易になったことが確認された。この設問は 2015 年度のアンケートからは削除された。

3.5 定期試験における得点率

リレーショナルデータモデル演習システムを用いた演習終了後およそ 2 ヶ月後、「データベース論」の成績評価のための定期試験が実施されている。この定期試験の結果から、リレーショナルデータモデルに関する問題の得点率を抽出したものを表 16 に示す。システム導入前後の比較のため、システム導入前の 2011 年度、2012 年度の 2 年分の得点率も示す。

リレーショナル代数の得点率は、システム導入前から 75% 程度の得点率があり、システムの導入による変化は確認できなかった。一方、正規化問題の得点率は、2013 年度

表 16 定期試験における得点率

学習項目	試験日/得点率				
	2011 年度	2012 年度	2013 年度	2014 年度	2015 年度
	2/7	2/5	2/4	2/10	2/9
リレーショナル代数	79%	78%	80%	75%	75%
正規化	46%	41%	59%	60%	63%

のシステムの導入以前よりも、10 ポイント以上向上していることが確認できる。本システムによる学習は、リレーショナル代数に比べ、正規化問題への学習に効果が高かったと考えられる。しかし、リレーショナル代数の得点率にはまだ及ばない。より学習効果を向上させる方策の検討が必要である。

4. むすび

本項では、リレーショナルデータモデル、とくに正規化理論の理解を目的に筆者らが開発したリレーショナルデータモデル演習システムの 3 年に及ぶ運用データと、改良の結果をまとめて報告した。

システムの導入により、定期試験における正規化問題の得点率向上が確認されている。

2013 年度に運用された初期のシステムでは、問題リレーションの関数従属性の不自然さが問題点として指摘された。そこで、2014 年度以降は関数従属性を考慮した値の組み合わせをもつテンプレートを用いることで、不自然さ（不完全さ）は解消されたが、完全さを求めると問題テンプレートの行数が増える問題点も残った。

今後は、さらに運用を続け、問題の不自然さを減少させつつ、学習の効果を向上させる改良を検討する。

参考文献

- [1] 長瀧寛之, 中野由章, 野部緑, 兼宗進: データベース操作の学習が可能なオンライン学習教材の提案, 情報処理学会論文誌, vol.55, no.1, pp2-15(2014).
- [2] 伊藤豪, 岡田信一郎: リレーショナルデータモデル演習システムの試作と評価, 2014 年電子情報通信学会総合大会, D-15-7(2014).
- [3] 山縣大輔, 岡田信一郎: リレーショナルデータモデル演習システムの正規化問題生成機能の改良と評価, 2015 年電子情報通信学会総合大会, D-15-10(2015).
- [4] 増永良文: リレーショナルデータベース入門 [改訂版], サ

イエンス社 (1991).