Larping the Past: Research Report on High-School Edu-Larp

Michał Mochocki

Abstract

The paper presents the results of "Live Action Role Plays in School Education," a quasi-experimental project piloted in 2010 and 2012 and conducted in 2013 by Games Research Association of Poland and educational publisher Nowa Era. In this project, larp was used in high-schools as a revision class before a test in History, the hypothesis being that larp as pre-test revision leads to higher retention of knowledge than traditional revision classes. Small sample size cannot be considered conclusive evidence, but the preliminary findings seem encouraging for further research. Beside students' test scores, the report provides non-standardized qualitative observational data on participant attitudes, plus information about practical issues of edu-larp research in public school environment. Next to the findings, the text reports on the research process and the rationale behind its modification.

Introduction

To be appreciated by formal education systems, larp needs efficiency in teaching the curricular content. This is to say that school teachers and administrators will be likely to reach for larp as an educational tool if they are convinced it can boost student achievements in core subjects such as, for example, History or Geography. An individual teacher, in turn, will focus on one particular core subject s/he teaches in a given class. Therefore, from the formal education perspective, the most important question is: how can larp be effective—or how effective can larp be—in teaching monodisciplinary subject-matter knowledge? The best way to find a reliable answer seems to be experimental research carried out in authentic school environment and focused on measurable learning outcomes.

Such was Live Action Role Plays in School Education, a quasi-experimental project piloted in 2010 and 2012 and conducted in 2013 by the Games Research Association of Poland and educational publisher Nowa Era (a local branch of Sanoma) in cooperation with volunteers from larp communities. In this project, larp was used as a revision class before a test, its functions being to recall, integrate, and consolidate knowledge from several previous lessons, which corresponded with one textbook unit. The initial hypothesis was that the use of larp as pre-test revision leads to higher retention of knowledge than traditional revision classes based on presentations and discussion.

The project aimed at verifying this hypothesis by quantitative data collected from standardized content-based tests administered in experimental (larp) and control (non-larp) groups, the results of which are presented in this paper. Additionally, the volunteers and teachers provided non-standardized qualitative observational data on participant attitudes, as well as the encountered problems and other practical issues they found worth mentioning. This data, also included here, will be valuable for teachers interested in bringing larp to the school environment. Finally, this paper reports on the research process and the rationale behind its modification.

Literature Review

The monodisciplinary and knowledge-centered approach to edu-larp fails to have its body of literature yet. Both theory and practice of edu-larping, as discussed by the international larp community in the Knutpunkt books since 2003 and the academic section of the Wyrd Con Companion Books since 2012, focus on the training of social skills, empathy, creativity, etc. These articles seem to dismiss larp as a tool for teaching monodisciplinary subject-matter; edu-larp expert Thomas Duus Henriksen said openly, "The method is awful for delivering hard knowledge."[1] Larp could be employed for teaching mixed curricular content in experimental game-based schools such as Østerskov Efterskole in Denmark[2] or inspire reflection on social

....

1. Thomas Duus Henriksen, "Learning by Fiction," in *As LARP Grows Up—Theory and Methods in LARP,* edited by Morten Gade et al. (Copenhagen: Projektgruppen KP03, 2003), 114. http://nordiclarp.org/w/images/c/c2/2003-As.Larp.Grows.Up.pdf

2. Malik Hyltoft, "Full-Time Edu-Larpers: Experiences from Østerskov," in *Playing the Learning Game: A Practical Introduction to Educational Roleplaying*, edited by Martin Eckoff Andresen (Oslo, Norway: Fantasiforbundet, 2012), 20-23.

issues such as discrimination and immigration,[3] but the prospect of larps used in regular schools seemed unlikely. The existing research shows that teachers prefer single-subject to cross-disciplinary activities[4] and are increasingly being forced "to 'teach to the tests' imposed by national standards programs."[5]

The 2013 Kickstarter campaign of Seekers Unlimited offers a solution: games designed "specifically to meet scholastic standards in math, science, social studies, language arts, etc." and supported with instruction "so any teacher can run them without us present."[6] However, this is an isolated initiative: Seekers claim they are probably the only company in the U.S. to use larps in this way.[7] Above all, data on edu-larp efficiency Seekers researchers have released concentrates on such indicators as intrinsic motivation or perceived competence,[8] while studies on the impact of larp on measurable scholastic achievement are yet to come. In general, the curriculum-based model of edu-larp is heavily under-researched, if not un-researched (see Bowman[9] in this book). A prior literature review did not find any experimental studies on the educational efficiency of such larps compared to traditional teaching, while such a study exists on tabletop role-playing games used to teach cellular biology to medical students in Brazil.[10] The 2013 Polish research project described in this article was an attempt to fill this gap.

Why should educators suspect larp could be an efficient tool for teaching curricular knowledge? As discussed at length in "Edu-Larp as Revision of Subject-Matter Knowledge," available in the *International Journal of Role Playing*,[11] edu-larp should be considered a type of Applied Drama/ Drama in Education as defined by Heathcote[12] or Bowell & Heap[13] building upon the theory and practice of drama educators. Edu-larp is also a *simulation* of socio-cultural environments according to Utne,[14] Larsson,[15] and Henriksen[16]; as a simulation, it is likely to improve student achievements, retention

....

3. Erik Aarebrot and Martin Nielsen, "Prisoner for a Day: Creating a Game Without Winners," in *Playing the Learning Game: A Practical Introduction to Educational Roleplaying*, edited by Martin Eckoff Andresen (Oslo, Norway: Fantasiforbundet, 2012), 24-29.

4. Sanne Harder, "Confessions of a Schoolteacher: Experiences with Role-playing in Education," in *Lifelike*, edited by Jesper Donnis, Morten Gade, and Line Thorup (Copenhagen, Denmark: Projektgruppen KP07), 234, http://nordiclarp.org/w/images/a/af/2007-Lifelike.pdf

5. Adam Blatner, "Creative Drama and Role Playing in Education," in *Interactive and Improvisational Drama*, edited by Adam Blatner (New York: iUniverse, 2007), Kindle edition, 1872.

6. Seekers Unlimited, "Creating Educational Live Action Role Playing Games" (Kickstarter.com, 2013), last accessed August 10, 2013, https://www.kickstarter.com/projects/160286787/creating-educational-live-action-role-playing-game/

7. Ibid.

8. Sarah Lynne Bowman and Anne Standiford, "Educational Larp in the Middle School Classroom: A Mixed Method Case Study," *International Journal of Role-playing* 5 (In press for 2015 publication); Sarah Lynne Bowman and Anne Standiford, "Edu-larp in the Middle School Classroom: A Qualitative Case Study," in DiGRA (Digital Games Research Association), Proceedings from the RPG Summit at DiGRA 2014. Salt Lake City, UT. 3-6 August 2014. (In press for 2014 publication).

9. Sarah Lynne Bowman, "Educational Live Action Role-playing Games: A Secondary Literature Review," in *The Wyrd Con Companion Book 2014,* edited by Sarah Lynne Bowman (Los Angeles, CA: Wyrd Con, 2014). (In this anthology).

....

10. Marco A. F. Randi and Hernandes F. de Carvalho, "Learning Through Role-Playing Games: an Approach for Active Learning and Teaching," *Revista Brasileira de Educação Médica* 37, no. (2013): 80-88. http://www.scielo.br/pdf/rbem/v37n1/12.pdf

11. Michał Mochocki, "Edu-*Larp* as Revision of Subject-Matter Knowledge," *International Journal of Role Playing* 4 (2013): 55-76. http://www.ijrp.subcultures.nl/wp-content/issue4/IJRPissue4mochocki.pdf

12. Dorothy Heathcote, "Contexts for Active Learning— Four Models to Forge Links Between Schooling and Society," Presented at NATD Conference 2002, accessed May 10, 2012, http://www.moeplanning.co.uk/wp-content/uploads/2008/05/dh-contexts-for-active-learning.pdf

13. Pamela Bowell and Brian S. Heap, *Planning Process Drama: Enriching Teaching and Learning* (London: Routledge, 2013), Kindle edition, 246-255.

14. Torstein Utne, "Live Action Role Playing—Teaching through Gaming," in *Dissecting LARP—Collected Papers for Knutepunkt*, edited by Petter Bøckman and Ragnhild Hutchison (Oslo: Knutepunkt, 2005), 24. http://nordiclarp.org/w/images/9/95/2005-Dissecting.Larp.pdf

15. Elge Larsson, "Participatory Education: What and Why," in *Beyond Role and Play: Tools, Toys and Theory for Harnessing the Imagination*, edited by Markus Montola and Jaakko Stenros (Helsinki: Ropecon ry, 2004.), 244. http://nordiclarp.org/w/images/8/84/2004-Beyond.Role.and.Play.pdf.

16. Thomas Duus Henriksen, "On the Transmutation of Educational Role-Play: A Critical Reframing to the Role-Play in Order to Meet the Educational Demands," in *Beyond Role and Play: Tools, Toys and Theory for Harnessing the Imagination*, edited by Markus Montola and Jaakko Stenros (Helsinki: Ropecon ry, 2004), 121, http://nordiclarp.org/w/images/8/84/2004-Beyond.Role.and.Play.pdf.

and understanding of complexity.[17] Because of its immediacy and emotional appeal, larp motivates students to *active involvement* according to Bowman[18] and Hyltoft,[19] which heightens its *educational potential*.[20] Active participation in a classroom activity may be encouraged by *collaboration*[21] or *competition*[22]; edu-larp combines both these incentives. In addition, larp allows for self-expression, while "the blending of competition,

cooperation, and self-expression," promises Karl M. Kapp, "will encourage the most engagement and activity with the gamified content."[23] To avoid failure when students face problems with immersion and acting, edu-larp can be *task-orientated* so that gameplay itself can keep the game running even in the absence of role-play; see the "mantle of the expert" approach in drama education,[24] "role distance… by the assumption of an attitudinal role,"[25] goal-orientated incentives,[26] or "productive tension."[27]

As I have previously discussed, to become eligible for mass implementation in schools, edu-larp should be 1) *monodisciplinary*: targeted at a single school subject, 2) *knowledge-oriented*: focused on knowledge and understanding of core, textbook subject-matter, and 3*) teacher-friendly*: eschewing time-consuming preparations, ideally through print-and-play form.[28] Furthermore, the edu-larp functions best as the "consolidator of knowledge"[29]; it should be applied not so much to introduce new content as to conduct a *final review* of previously learned things. As experts claim, educational content should be embedded in the game, making it part of gameplay[30] and the game itself should be "embedded in a program of instruction" rather than as a "stand-alone simulation."[31] Hence, in this project, larp was positioned as a pre-test revision class at the end of a large content unit covering several weeks of class work in a high school course in History.

…………………………….

17. Thomas Duus Henriksen, "Moving Educational Role-Play Beyond Entertainment," *Teoria de La Educacion en la Sociedad de la Informacion* 11, no. 3 (2010): 241, http://es.youscribe.com/catalogue/libros/conocimientos/ciencias-humanas-y-sociales/moving-educational-role-play-beyond-entertainment-1781104/; David Crookall, Rebecca Oxford, and Danny Saunders, "Towards a Reconceptualization of Simulation: From Representation to Reality," *Journal of SAGSET (Society for the Advancement of Games and Simulations in Education and Training* 17, no. 4 (1987): 150; Lisa Galarneu, "Authentic Learning Experiences Through Play: Games, Simulations and the Construction of Knowledge," in Changing Views: Worlds in Play. Proceedings of DiGRA 2005 Conference, 2005, 4, http://www.digra.org/dl/db/06276.47486.pdf

18. Sarah Lynne Bowman, *The Functions of Role Playing Games. How Participants Create Community, Solve Problems and Explore Identity* (Jefferson: McFarland & Co, 2010), 86.

19. Malik Hyltoft, "Four Reasons Why Edu-Larp Works," in *LARP: Einblicke. Aufsatzsammlung zum MittelPunkt 2010*, edited by Karsten Dombrowski (Braunschweig: Zauberfeder, 2010), 56, http://www.zauberfeder-verlag.de/Produkte/MP10/MP10_Artikel-04.pdf

20. Myriel Balzer, "Immersion as a Prerequisite of the Didactical Potential of Role-Playing," *International Journal of Role-Playing*, Issue 2 (2011): 35, http://www.ijrp.subcultures.nl/wp-content/issue2/IJRPissue2-Article3.pdf; Paul Heyward, "Emotional Engagement Through Drama: Strategies to Assist Learning through Role-Play," *International Journal of Teaching and Learning in Higher Education* 22, no. 2 (2010): 198, http://www.isetl.org/ijtlhe/pdf/IJTLHE751.pdf; Yaraslau Kot, "Educational Larp: Topics for Consideration," in *Wyrd Con Companion Book*, edited by Aaron Vanek and Sarah Lynne Bowman (Los Angeles, CA: Wyrd Con, 2012), 122; Frida Sofie Jansen, "Larp the Swedish Way: An Educational Space Odyssey," in *Playing the Learning Game: A Practical Introduction to Educational Roleplaying*, edited by Martin Eckoff Andresen, 30-35 (Oslo, Norway: Fantasiforbundet, 2012), 32; Bowell and Heap, *Planning Process Drama*, 613-614.

21. Eyal Szewkis et al., "Collaboration Between Large Groups in the Classroom," *International Journal of Computer Supported Collaborative Learning* 6, no. 4 (December 2011): 566; Marco Villalta et al., "Design Guidelines for Classroom Multiplayer Presential Games (CMPG)," *Computers & Education*, 57, no. 3 (November 2011): 2045.

22. Crookall, Oxford and Saunders, 165; Byron Reeves and J. Leighton Read, *Total Engagement* (Boston: Harvard Business Press, 2009), Kindle edition, 28-29.

…………………………….

23. Karl M. Kapp, *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education* (John Wiley and Sons, 2012), Kindle edition, 3471-3472.

24. Gavin Bolton, "A conceptual framework for classroom acting behavior" (Ph.D. diss., Durham University, 1997), 387, http://etheses.dur.ac.uk/1627/

25. John Carroll and David Cameron, "Playing the Game, Role Distance and Digital Performance," *Applied Theatre Researcher* 6, no. 1 (2005): 11, http://www.griffith.edu.au/__data/assets/pdf_file/0009/54945/playing-game.pdf

26. Thomas Duus Henriksen, "Moving," 249.

27. Dorothy Heathcote, "Productive Tension. A Keystone in 'Mantle of the Expert' Style of Teaching," *The Journal for Drama in Education* 26, no. 1: (2010), 8-23. http://www.mantleoftheexpert.com/wp-content/uploads/2010/03/Jan-101.pdf

28. Mochocki, "Edu-*Larp.*"

29. Steve Guscott, "Role play as active history," in *A Practical Guide to Teaching History in the Secondary School*, edited by Martin Hunt (New York: Routledge, 2007), 40.

30. Alejandro Echeverría et al., "A Framework for the Design and Integration of Collaborative Classroom Games," *Computers & Education* 57, no. 1 (August 2011): 112.

31. Kapp*, The Gamification*, 2449-2453.

METHODS

Participants

The main stage of research was launched in 2012/2013 as a joint project by Games Research Association of Poland and the Nowa Era publishing house, with the support of volunteers from larp communities in Gda sk and Kraków. Responsibilities were divided as follows:

GRAP: The overall design of experiment and the preparation of all written materials: larp scenario, instruction for gamemasters, and standardized tests.

Nowa Era: Recruitment of schools and teachers for the project in two cities: seven schools in Kraków and six in Gdańsk. The choice of Kraków and Gdańsk resulted from the availability of volunteers; the selection of particular schools was up to the recruiters.

Volunteers: Running the larps in schools according to the scenario delivered by GRAP and authored by myself. In Kraków, the partner was Pospolite Ruszenie Szlachty Ziemi Krakowskiej (PRSZK), a society for historical reenactment and education, well-experienced in school larps after the DEMOcracy Project.[32] The Kraków games were coordinated by Łukasz Wrona, Vice President of PRSZK and the former head of the DEMOcracy Project. In Gdańsk, the team comprised experienced larpmasters from the Science Fiction/Fantasy club Zardzewiały Topór, the Science Fiction Club of Gdańsk, or non-affiliated. Gdańsk was coordinated by Krzysztof Chmielewski, the editor of the first Polish book on larp (*Larp. Myśli i szkice*, 2012), counted among the best larpwrights in Poland.

Teacher participation: While the larp class was conducted by 2-3 volunteers, the teacher was present as a passive observer. The teacher's primary contributions to the project was to distribute and grade tests on schedule and to send test scores to the project coordinator. The complete measurement in the given group would entail two post-larp tests and two non-larp tests, as discussed below.

Students: All groups were in the second year of high school (aged 17-18), comparable to 12th grade in K-12, although Poland requires another, final high-school grade attended by 18-19-year-old. The total

....

32. Michał Mochocki and Łukasz Wrona, "DEMOcracy Project. Larp in Civic Education," in KoLa 2013. Larp Conference Proceedings, edited by Jakub Tabisz (Wrocław: Wielosfer, 2013), 20-27. http://issuu.com/wielosfer/docs/kola13-publication/

number of participants was 132 in the 2010 pilot study, 105 in the 2012 pilot study, and 153 in the main 2013 research.

Legal issues: Innovations and experiments in Polish public schools are regulated by an April 9, 2002 decree by the Minister of Education and Sport.[33] The decision to introduce and test innovative methods in a school is within the powers of the principal and teachers' council, and should be consulted with the parents' council.[34] With approval from the parents' council, it is also within the principal's abilities to start cooperation with NGOs and other organizations as specified by the 1991 Act on Education System.[35] Here, the only third party intervention was the larp itself, conducted by external volunteers, but still supervised by the local teacher, who incorporated it in the ongoing didactic process as pre-test revision. Tests were administered, were marked by the teacher as part of regular grading and diagnostic practice; and the scores were anonymized (e.g. "5 students got 12 points; 6 students got 11 points"). In no way were the researcher or volunteers given access to students' personal data. None of the above required individual consent forms from parents or students according to Polish research policies.

Apparatus

According to Mayer, "a richer vision of how to improve educational practice emerges through a multileveled set of measures, which may include both quantitative and qualitative measures."[36] In this project, quantitative data should help answer the question of larp's efficiency in building the knowledge of core subject matter. Qualitative data brings insight into the practical side of organizing larps in school and also into the affordances and constraints of conducting

....

33. MENiS, Rozporządzenie Ministra Edukacji Narodowej i Sportu z dnia 9 kwietnia 2002 r. w sprawie warunków prowadzenia działalności innowacyjnej i eksperymentalnej przez publiczne szkoły i placówki (Dz. U. z 2002, Nr 56, poz. 506) (Decree of Minister of National Education and Sport on the requirements for innovative and experimental activity in public schools and facilities, April 9, 2002), http://www.dziennikustaw.gov.pl/DU/2002/s/56/506/1/

34. Ibid., §4.

35. *Ustawa z dnia 7 września 1991 o systemie oświaty* (Dz.U. z 2004, Nr 256, poz. 2572) (Act on Education System, 7 September 1991), Chapter 4, Art. 56, http://www.dziennikustaw.gov.pl/DU/2004/s/256/2572/1/

36. Richard E. Mayer, "The Failure of Educational Research to Impact Educational Practice: Six Obstacles to Educational Reform," in *Empirical Methods for Evaluating Educational Interventions*, edited by Gary D. Phye, Daniel H. Robinson and Joel Levin (San Diego: Elsevier Academic Press, 2005), 77.

pedagogical research on larp.

a) Quantitative Assessment

Unlike in America, "standardized test" here does not mean "based on U.S. Common Core standards," but rather "designed according to a precise, measurable, and replicable standard." Learning outcomes were measured with standardized tests with 12 closed questions. Students were to provide the right answer constituted by a short and specific piece of knowledge, e. g. "Which country was the most interested in getting the Polish-Lithuanian Commonwealth into war against Turkey? Answer: Austria." Standardized tests came as a convenient choice; as Łobocki points out, they a) face all students with identical challenges, b) allow for fair and measurable assessment with minimized subjectivity, c) allow for simultaneous assessment of large groups, d) make it possible to compare the results of different groups and schools.[37]

Both the experimental and control groups were taught the same content by their history teacher. The follow-up test was based strictly on the textbook unit that the students were supposed to learn. All tests were designed by the research coordinator (Mochocki) and approved by the teacher as a valid tool of measurement.

All tests were administered twice, at short and long distance. In other words, the test was announced, took place, and was graded for the first time about 7 days after the revision. The same test was administered again about 30 days after the revision, this time unannounced and ungraded. In the pilot studies, we only administered one test, the same for both experimental (larp) and control (non-larp) groups. In the 2013 main stage, however, each group received two separate tests for different textbook units: one test was experimental (larp), the other control (non-larp). The group's achievements in the experimental test(s) were to be compared with control test(s). This was a design challenge: to make comparison possible, the difficulty level of both tests should be equivalent, even though they covered different content.

All experimental (larp) tests were based on the "Poland in 18th century" unit, with the control (non-larp) tests from another unit (see Table 1, where the test covers Europe in the first half of 19th century).

[38] To establish the desired equivalence, each standardized test for the control (non-larp) stage was designed as an equivalent counterpart of the "normative" test on 18th century Poland (larp). The equivalence of the tests as a whole was based on the equivalence of all its 12 items, considered on a one-on-one basis. In other words, every question in the control (non-larp) test was paired with one corresponding question in the experimental (larp) test, and carefully crafted as equivalent in format, topic, level of detail, and relative importance/prominence of the content item. For example, if one test asks for the name of a well-known military leader counted among prominent public figures of the period, so does the second test in relation to the other period/country. The difficulty level of both tests was evaluated by the teachers—who are treated as subject-matter experts[39]—and approved by them as equivalent.

Table 1: Sample Equivalence of Test Content

Test 1 (Poland in 18th century)	Test 2 (the first half of the 19th century)
1. What was the name of a "ministry" of education created in 1773? (Answer: Commission of National Education)	1. Which act replaced the constitution of the Kingdom of Poland in 1832 (Answer: Organic Statute)
2. The First Partition of Poland was carried out as "punishment" for [what event?] (Answer: the Bar Confederation)	2. The Congress of Vienna was summoned in the aftermath of [what event?] (Answer: the defeat of Napoleon / the collapse of the Napoleonic rule)

The collection of this data was entirely in the hands of the teachers, who were asked to administer four tests on schedule: 1 week and then 1 month after the revision class, in both the experimental and the control group. The teachers were also asked to check the results on a 12-point scale, cross out students who did not participate in the larp, and send anonymized scores to the researcher.

It was important that the first test be taken about

....

37. Mieczysław Łobocki, *Metody i techniki bad pedagogicznych* (Kraków: Oficyna Wydawnicza "Impuls," 2003), 150.

....

38. Please note: it does not have to be the next textbook unit after the larped one. The length of time between the larped and non-larped units seems to be of little relevance. In the 2013 study, all groups took *Facing the Second Partition* (Poland in 18th century) as the larped (experimental) class, but different textbook units were used in control groups.

39. D'Agostino, "Measuring," 142.

7 days after the revision class; if it was delayed for another week or two (e. g. due to the teacher's illness), it could not be used as measurement of short-term retention. The tight schedule made it is a fairly demanding task, and not all teachers were able to meet it—as clearly demonstrated in the below Results by the number of schools that did not provide complete (or any) test scores.

a) Qualitative Assessment

The qualitative observational data—i.e. reports of specific issues observed in classroom practice and in cooperation with the school as institution—was collected primarily for the needs of larp/drama educators and teachers interested in the practical side of bringing edu-larp to schools. Delivered by larpmasters—in this case, volunteers who administered the larps in class—and teachers in written reports and online/telephone interviews, this data comes from all 13 games in the 2013 run and 10 games from the pilot studies in 2010 and 2012. The volunteers were asked to report on:

- students' and teachers' attitude to larp
- ratio and level of (un)cooperativeness
- technical and organizational problems
- any other factors that they found to have influenced the larp or the larp experience in a positive or negative way

This data was gathered in a non-standardized form as paper notes, e-mails, notes from telephone conversations and face-to-face meetings. It was then processed in the form of listings and descriptions, leading to the report in the Results below.

Procedure

With non-random assignment of students to groups, as the researchers tested already-formed classes, the project constituted a quasi-experimental study, or "design research" as distinguished from strict experimental research by Collins et al.[40] The project stayed in line with the idea of design research: "the design is constantly revised based on experience, until all the bugs are worked out."[41]

....

40. Alan Collins, Diana Joseph and Katerine Bielaczyc, "Design Research: Theoretical and Methodological Issues," *Journal of the Learning Sciences* 13, no. 1 (2004): 15-42. http://treeves.coe.uga.edu/EDIT9990/Collins2004.pdf

41. Ibid, 18.

Overall, three methods of experimental pedagogical interventions have been considered and tested within 3 years of pilot studies and the main study. As described by Valentine & Cooper (2005) in their summary of William McCall:

a *one-group study* often compares students before and after receiving some intervention. An example of a one-group design would be if a researcher noted the achievement levels of students in a school both before and after a uniform policy was introduced. If the students' achievement changed, the researcher may be tempted to say that it was due to the policy. If their achievement stayed the same, the researcher may conclude that the policy had no effect…

an *equivalent groups study* assesses the effect of an intervention by comparing a group of students who received the intervention with a group of students who did not receive the intervention…

…the *rotation* or crossover design… is that participants receive one level of an intervention and then another. To borrow McCall's example, if a teacher were interested in the effects of praising versus scolding students, he or she could first praise a student and note the effects. The teacher could then scold the student and note the effects. This could be done with several students. If implemented in conjunction with other features (e.g., using random assignment to determine which students receive which condition first), the crossover design presents a very strong basis for making inferences about an intervention's effectiveness.[42]

This study hypothesized that larp as pre-test revision leads to higher retention of knowledge than traditional revision class based on presentations and discussion. In operational terms, the higher retention of knowledge would be demonstrated if tests covering the larped units show significantly higher scores than tests from the non-larped units. More specifically, as test-based measurement breaks down in two components—short-distance (1 week) and long-distance (1 month)—it is possible to hypothesize that larped units should have *slightly* higher scores in short-distance tests and *significantly* higher scores in long-distance tests. Why? The first test had been

....

42. Jeffrey C. Valentine and Harris M. Cooper, "Can We Measure the Quality of Causal Research in Education?," in *Empirical Methods for Evaluating Educational Interventions*, edited by Gary D. Phye, Daniel H. Robinson, and Joel Levin (San Diego: Elsevier Academic Press, 2005), 90-91.

announced and graded, with students in both groups expected to do pre-test studying at home. Therefore, the score could be much more influenced by last-minute cramming than by the form of in-class revision. This problem is universal in ludic edu-innovations. In graded evaluations, they usually do not show better results than traditional methods; they only shine in measurements of "natural" long-term retention where students make no special pre-exam preparation.[43] For this reason, the long-distance tests—unannounced and non-graded—were selected as the primary means of verification of the main hypothesis.

a) Procedure in 2013 Study

With the abandonment of equivalent groups in favor of the one-group method after the pilot studies (see below), in the main 2013 study, the total procedure followed six steps in two stages in all the participating groups.

During the experimental stage:

- a larp was played in class as pre-test revision of a single textbook unit covering 18th-century Poland;
- 1 week later, students took an announced and graded standardized test;
- 3 weeks later, they retook the same test with no prior announcement.

During the control stage:

- another textbook unit, Poland and Europe in the first half of 19th century, was completed with a typical pre-test revision led by the teacher and mostly based on presentation and discussion;
- 1 week later, students took an announced and graded standardized test;
- 3 weeks later, they retook the same test with no prior announcement.

Turning to how this experiment was positioned in the ongoing process of instruction in schools:

1. A given unit from the History textbook was discussed in class over several weeks for one 45-min. period a week (two in humanities-focused programmes), followed by home assignments. Methods of instruction varied; predominantly, they comprised presentation, discussion, Q&A, worksheets, source text analysis, etc., with the strong control of the teacher. Each new lesson covered a new part of the overall content of the unit;

2. A pre-test revision class reviewed the previously discussed text/commissioned reading. Typically, it was still based on content presentation, Q&A, and group discussion. In the control groups/stages, teachers conducted it in the traditional way. In the experimental groups/stages, the revision class took the form of larp.

3. A formal test measured the learning outcomes based on the given unit. It had been announced in advance and was graded by the teacher. This was the short-distance test, administered 1 week after the larped (experimental) or non-larped (control) revision class.

Normally, a student who passes the test no longer needs to return to the content of a completed unit unless they choose to take History as a component of the maturity exam (comparable to A-level) at the end of high-school. In the project, both experimental and control groups took the same test again 3 weeks later (4 weeks after the revision), without prior notice.

a) Procedure in 2010 and 2012 Pilots

Initially, the idea was to use the *rotation method* based on the comparison of test scores of experimental groups (classes in which larp was used as pre-test revision of a book unit) and control groups (classes revising the same book unit for the same test with no larp).[44] In line with the rotation method, recommended as the most reliable by experts on pedagogical research,[45] in one semester the experimental group would play a larp while the control group would not and in the next semester they would switch; the former control (non-larp) group would now play another larp as the experimental group and vice versa.

With this goal in mind, a pilot study in 2010 included 3 larps (addressed below as A, B and C) on History classes in two high schools in Bydgoszcz, Poland. The researchers obtained quantitative data (test scores) from 132 students in total, as well as

....

43. Teresa Siek-Piskozub, *Gry, zabawy i symulacje w procesie glottodydaktycznym* (Poznań: Wydawnictwo Naukowe UAM, 1995), 46.

....

44. The concept for this research was first presented in: Michał Mochocki, "Teatralne gry fabularne (LARP-y) w nauczaniu szkolnym," *Homo Ludens* 1/2009, 177-189.

45. Władysław Zaczyński, *Praca badawcza nauczyciela* (Warszawa: Wydawnictwa Szkolne i Pedagogiczne, 1995), 99-100; Łobocki, Metody, 112-114; Albert W. Maszke, Metody i techniki badań pedagogicznych (Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego, 2008), 184-185.

non-quantitative data from observation and reports (see Qualitative Assessment, above). Additionally, 6 other editions of the same larps run by myself or my co-workers in high schools in Bydgoszcz, Gdańsk, and Myślenice brought only qualitative information, as no tests were delivered by the teachers. In the second semester, the procedure was supposed to be repeated with group rotation, to reduce the risk of outcomes being distorted by "pre-instructional student differences in aptitude or achievement."[46] However, not even once did this procedure occur as detailed under Discussion below. Without the rotation, it effectively turned into the *equivalent-groups method*. The next pilot study—conducted in 2012 by teacher and researcher Jerzy Szeja in a high-school in Łochów near Warsaw,—used the equivalent groups deliberately with no attempt at rotation to make it consistent with the 2010 study. However, my growing distrust in the credibility of this method resulted in the change of procedure for the 2013 research. The new procedure reached its final shape in the six steps listed above. Further justification for the selection of one-group method is detailed in the Discussion below.

RESULTS

a) Qualitative

The following problems were observed at least once in the 23 games played in Polish high schools in 2010, 2012, and 2013.

"*Swarm tactics*," a term by larpmaster Tomasz Łomnicki: In the middle phase (unofficial negotiations of all parties; see Appendix—Game Materials), students do not break into small groups but form one chaotic and noisy crowd, ignoring the fact that their opponents can eavesdrop. Tested solution: specifically warn against this problem in the briefing.

Small rooms: In a small classroom, it is difficult to accommodate the first two phases (prepping in separate factions, then unofficial talks in small groups). Obvious solution: use the corridors. However, loud talk in the corridors irritates teachers working in other classrooms.

Refusal to role-play: Some students are unwilling to get involved in role play, especially those with no experience with drama or RPG. A student may openly refuse to play, or use passive resistance[47]: watch

and follow the herd without active involvement. The rate of students who remain uninvolved and passive throughout the game is estimated by larpmasters at about 10% (2-3 people in a class). Tested solutions: approach the student as out-of-game larpmaster and a) make the student feel his/her character is very important, which works with ambitious and easily bored students, b) ignore role-play by encouraging them to simply focus on task completion, which works with shy and self-conscious ones, c) focus their attention on a personal conflict/competition with another character, which works with competitive and aggressive students.

Game-based logic:[48] Students may approach the game as a "puzzle" with specific victory conditions and a pre-defined outcome. These individuals feel frustrated when they cannot find the right combination. Solution: Make it clear in the briefing that the larp is a simulation of complex socio-political reality with no obvious answer to be found.

Anachronisms: Students may use references to events or technology unknown in the period, ruining the potential for immersion and reasonable debate. Tested solution: specifically warn against anachronisms in the briefing by explaining that using anachronisms generates "system errors," as there is no way other characters can logically respond. Without this warning, students feel tempted to test the limits, but when it is made clear that anachronisms disrupt the game and can easily ruin it, they behave responsibly.

Teacher's anxiety: The teacher may feel stressed by the prospect of a third party researcher studying the achievements of his/her group in comparison with other schools. One local coordinator suspects that some of the teachers deliberately "sabotaged" the test schedule due to this anxiety. It could be the main reason why so many teachers did not deliver test results (see Results and Findings below).

The most important positive findings were:

Positive attitude of students: The great majority of students seemed to enjoy the larp class, and declared they were willing to repeat this experience. Krzysztof Chmielewski, the local coordinator in Gdańsk, reports that the edu-larp inspired several students to take on larp as a hobby and one of them started to write larps herself.

Positive attitude of teachers: Also, the majority of teachers found larp a valuable tool and expressed interest in further cooperation with the volunteers.

Since the qualitative data deliberately focused on problems, the negative findings outlined here

....

46. Joel Levin, "Randomized Classroom Trials on Trial," in *Empirical Methods for Evaluating Educational Interventions*, edited by Gary D. Phye, Daniel H. Robinson, and Joel Levin (San Diego: Elsevier Academic Press, 2005), 7.

47. Jansen, "Larp," 33.

....

48. J. Tuomas Harviainen, Timo Lainema, and Eeli Saarinen, "Player-Reported Impediments to Game-based Learning," in Proceedings of 2012 DiGRA Nordic, 2012, http://www.digra.org/dl/db/12168.02279.pdf

outnumber the positive findings. However, the overall conclusion falls strongly on the positive side (see Discussion / Findings below).

a) Quantitative: Pilots

In each case, the students' achievements (test scores) were based on standardized 12-item tests with closed questions (see above under Apparatus).

Table 1 covers the 2010 pilot study with 3 larps (addressed below as A, B and C) on history classes in two high schools in Bydgoszcz obtaining test scores from 132 students in total. Each larp group was paired with a control group and they were supposed to switch roles in the next semester which did not happen (see the above section regarding the inadvertent change from the rotation method to equivalent-groups). The average score (%) was calculated for the entire group: the sum of all individual scores was divided by the number of students and converted to a percentage.

Ultimately, in 2010, the 3 experimental (larp) groups outperformed the 3 control groups in all tests, both in short (1 week) and long (1 month) distance.

Table 2 shows the results of the 2012 pilot study by Jerzy Szeja in Łochów with 105 students in total. Test scores of the experimental group (larp group D) are compared with the score of 3 control groups that took a traditional, teacher-led revision. The average score (%) is calculated as above: the sum of all individual scores is divided by the number of students and then converted to %.

Table 2: Average Score in 2010 Pilot Study (equivalent-groups method)

Group	1 week after revision (ca. 7 days after larp/revision)	1 month after revision (ca. 30 days after)
Larp group A	75.5% (30 students)	81.6% (29 students)
Control group A	— *	50% (24)
Larp group B	66.8% (28)	60.2% (27)
Control group B	25.8% (21)	22.9% (22)
Larp group C	72.5/85% (26/23)**	52.5% (25)
Control group C	45% (28)	36% (22)

*   No parallel test was carried in due time here, so there are no results of the control group.

**   3 tests came back unsigned and with no answers at all. If included in the total score, it is 72.5%; if excluded, the score is 85%.

Table 3: Average Score in 2012 Pilot Study (equivalent-groups method)

Group	1 week after revision	1 month after revision
Larp group D	69.1% (24)	73.4% (21)
Control group D-1	70.7% (25)	65.3% (24)
Control group D-2	65.9% (23)	— (not delivered)
Control group D-3	72.22% (33)	90.5% (28)

In the short-distance test, all groups have similar average results (65.9 to 72.2%). The long-distance retest shows a decrease in the score of control group D-1, but an increase for control group D-3 and for the larp group D. Class D-3 achieved top score in both tests. According to their history teacher, this ideally illustrates the difference in the overall competence between the groups: the larp class and control classes D-1 and D-2 are on a comparable level, while control D-3 has always been best-performing. This information was obtained by Jerzy Szeja, the researcher conducting the 2012 pilot study and fellow teacher in the school.

a) Quantitative: 2013 Research

6 out of 13 schools participating in the 2013 edition of the project had to be removed from the pool of quantitative data because of delayed or incomplete testing in the first experimental stage.

Table 3 presents the results of the remaining 7 groups (153 students in total), delivered in varying degrees of completeness. The "post-larp" (experimental) and "non-larp" (control) sections show the average scores in tests following the larp and in the non-larp revision class, respectively, which were conducted in the short- (ca. 7 days) and long-term (ca. 30 days). In each case, the table shows:

1. Point value of the average score of the group: the sum of individual scores divided by the number of students. With 12 questions worth 1 point each, the maximum score is 12 and the minimum is 0.
2. (in parentheses) Percentage of the above-mentioned point value with 12 as 100%.
3. [in brackets] Number of tests taken into account.

Out of the seven schools, only 1 (Group 4, in **bold**) provided the full set of data for all four tests, including the second (non-larped, control) stage, hence the empty boxes in the table.

Table 4. Average Scores in 2013 Research (one-group method)

Group	Post-larp tests (Experimental)		Non-larp tests (Control)	
	1 week	1 month	1 week	1 month
Group 1	9.43 (78.6%) [16 students]	8.47 (70.58%) [17]	7.19 (59.91%) [21]	
Group 2	9.36 (78%) [25]	10.25 (85.41%) [24]	8.36 (69.66%) [19]	
Group 3	10.90 (90.83%) [32]	11.65 (97.08%) [29]		
Group 4	**8.43 (70.25%) [23]**	**10.52 (87.66%) [19]**	**9.65 (80%) [23]**	**9.04 (75%) [25]**
Group 5	8.9 (74.16%) [22]	8.47 (70.58%) [21]		
Group 6	7.06 (58.83%) [15]	6.58 (54.83%) [12]		
Group 7	8.22 (68.5%) [20]	6.36 (53%) [19]		

No division into Ctrl and Exp groups here. Each group first underwent an experimental (larped) revision class before a test, and then a control (traditional) revision before another test.

DISCUSSION

a) Findings

As discussed in Methods, the primary hypothesis—larp as pre-test revision leads to higher retention of knowledge than traditional revision class based on presentations and discussion—would be verified if the experimental (larp) groups had achieved significantly higher scores in the unannounced long-distance tests. This hypothesis indeed happened in the only complete data set (Group 4) in the 2013 study: the short-distance post-larp score (70.25%) was worse than non-larp (80%), but in the long distance, the post-larp score rose from 70.25% to 87.66% with the non-larp score dropping from 80% to 75%. Should this pattern consistently repeat in a large sample, the hypothesis would be verified. At this moment, the small sample (~23 students) only counts as preliminary findings that encourage further research.

The secondary operational hypothesis claimed that experimental (larp) groups should have slightly higher scores in short-distance tests. For this hypothesis, the comparison can be made in three 2013 groups. For Groups 1 and 2, the average score in the short-distance (1 week) post-larp test is higher than in the corresponding non-larp tests, the difference being remarkable in Group 1, with 78.6% vs. 59.91%, and less so in Group 2, with 78% vs. 69.66%. However, in Group 4, the scores show the opposite: it is the post-larp test that has lower short-distance performance: 70.25% vs. 80%. Again, only a much larger sample could bring conclusive evidence.

Because of the change of method from equivalent-groups in 2010/2012 to one-group in 2013, it does not make sense to put them in one box. They should rather be treated as separate data sets.

In the pilot study of 2010, test scores from all three experimental larp groups A-C are significantly higher than in all three control groups A-C, both in short- and long-distance. In the pilot of 2012, the one and only experimental larp group D performed slightly better than control groups D-1 and D-2 but worse than control group D-3. In both pilot studies (A-D) put together, the larp scores were remarkably higher than non-larp scores 3 times; on 2 occasions, the numbers were more or less equal, and a non-larp group performed better 1 time.

How reliable are these numbers? Contrary to the initial plan, the pilot results were obtained with the equivalent-groups method without group rotation. "The problem with the equivalent group's design… is that we cannot be sure that the groups are truly equivalent,"[49] as said by Maszke[50] and Zaczyński.[51] Indeed, according to their teacher, the non-larp group D-3 which topped the score in 2012 had always been best-performing. In research on already-formed school classes, when researchers have no means of improving equivalence, it is clear that groups will often not be equivalent. In each particular case, the difference in tests scores might be attributed to initial competence levels.

A stronger case for the usefulness of larp is found not in the one-to-one comparisons, but in the total average scores (see Tab. 5) of all 2010-2013 studies.

Table 5 (next page) puts together all scores from the pilot and post-pilot studies and calculates the average score of all larp-based tests compared to the collective average of the non-larp ones.

Larp-based revisions in 11 classes [258 students] correlate with the average of ~70% score on short-distance test, with only 1 out of 11 dropping below 60%; while the corresponding short-distance average for the 8 non-larp classes [200 students] is ~65%, with 3 dropping below 60%. In the long-distance retest, the 11 larp classes [243 students] retain the average of ~70%, with 3 out of 11 below 60% and none below 50%, while the average of the 6 non-larp ones [145 students] is ~55%, 3 of them below 50%.

In these figures, all experimental groups are counted as one large meta-group, as are control groups. Therefore, the non-equivalence between two particular small groups is less worrying. There is still potential non-equivalence between these combined meta-groups; the experimental meta-group may have more top-performing and/or less bottom-performing students than the control meta-group or vice versa. Still, with 258 "experimental" students and 145 "control" ones, it may be believed that the ratio of top- and bottom-performers in both meta-groups is statistically similar.

All these findings should be approached with caution; the preliminary results suggest—not prove—the superiority of larp as a revision class. They are not conclusive enough to reject the null hypothesis, i.e. that larp as a revision class may have no significant effect on test performance. One thing, however, is certain: not even once did larp seem to negatively impact the group's test performance. Further experiments should not meet objections.

When it comes to the qualitative findings, it should be concluded that edu-larp is widely applicable in regular school environments. This was

....

49. Valentine and Cooper, "Can We Measure," 91.

50. Maszke, *Metody*, 184.

51. Zaczyński, *Praca*, 97.

Table 5: Total Average Scores of Post-Larp and Non-Larp Tests

	Post-larp 1 week	Post-larp 1 month	Non-larp 1 week	Non-larp 1 month
Pilot 2010*	[81 students]	[81]	[56]	[68]
A	75.5%	81.6%	—	50%
B	66.8%	60.2%	25.8%	22.9%
C	85%	52.5%	45%	36%
Pilot 2012*	[24 students]	[21]	[81]	[52]
D	69.1%	73.4%	70.7%	65.3%
D			65.9%	—
D			72.22%	90.5%
Project 2013*	[153 students]	[141]	[63]	[25]
Group 1	78.6%	70.58%	59.91%	—
Group 2	78%	85.41%	69.66%	—
Group 3	90.83%	97.08%	—	—
Group 4	70.25%	87.66%	80%	75%
Group 5	74.16%	70.58%	—	—
Group 6	58.83%	54.83%	—	—
Group 7	68.5%	53%	—	—
Total Average	**74.64% [258 students]**	**74.16% [243]**	**63.64% [200]**	**53.66% [145]**

*   *It should be remembered that Pilots 2010 and 2012 used equivalent-groups, i.e. the "post-larp" (experimental) group is different than the "non-larp" (control) group. By contrast, in Project 2013 there are "post-larp" and "non-larp" scores the same group that went through the experimental and the control stage.*

confirmed not only by the 23 larps carried out in high schools as part of this project and pilot trials, but also by 56 first-level larps run in junior high schools under the DEMOcracy Project.[52] Out of the 79 games, there were only two reports of serious struggles with group's uncooperativeness when the larpmasters failed to get the game running because of the collective refusal to participate.

The usual 10% of people who, according to the qualitative reports, stay uninvolved and passive while others are playing is unfortunate. Still, it is very unlikely that other alternatives, e. g. a traditional pre-test revision class, would be more successful. Student disengagement is a huge problem[53] in general and all school activities tend to suffer from it. If edu-larp is reported to keep 90% students actively involved for over an hour, this is a strength—not weakness.

Generally, with the exception of teacher's failure to deliver test scores, all reported problems were observed rarely and can be met with field-tested solutions.

b) Limitations

The overall idea of standardized tests is very practical: well-established in the school environment and, thus, easily accepted by teachers and students. However, should the test be repeated after 3 weeks with the very same questions? It is clear that students who received correct answers in teacher's feedback after test A are better prepared for test B.[54] This could be seen as a serious limitation of the study (see the discussion of the AB method and unacknowledged interdependence between A and B in Levin).[55] However, the object of study was long-term retention of this particular knowledge that had been measured in the first test, in which case a different set of

....

52. Mochocki and Wrona, *DEMOcracy*.

53. Helen M. Marks, "Student Engagement in Instructional Activity: Patterns in the Elementary, Middle, and High School Years," *American Educational Research Journal* 37, no. 1 (Spring 2000): 154.

....

54. D'Agostino, "Measuring," 118.

55. Levin, "Randomized," 11.

questions does not seem a viable option. The fact that students had been given the answers makes it even more relevant to the school practice. This is the regular cycle: a) teach, b) revise, c) test, d) give feedback on test results with correct answers in the hopes that e) students will long retain some of this knowledge. The research hypothesis can now be rephrased as: if the b) "revise" stage takes the form of edu-larp, then e) long-term retention will be higher. Thus, the purpose should be to check how much knowledge stays with students after they have been tested and graded. Retaking the same test unannounced 3 weeks later should then be counted among "measurement methods… that match the purposes of the intervention under study."[56]

The weakest spot of the 2013 study is the lack of researcher's control over the process of testing, which was left entirely in the hands of teachers. It was the teacher's responsibility to prevent cheating on the tests, to meet the test schedule, to eliminate non-larpers from the data pool, and to deliver the exact scores without miscalculations or typos. Each teacher was testing only one group to avoid work overload, but the possibility of human error remains nevertheless. Still, the biggest problem with teacher's responsibility was not errors, but the failure to conduct the tests and deliver the scores. With 13 schools in the project, 7 delivered no quantitative data, 5 did it partially (with very limited usefulness of this data), and only 1 provided the full set (Group 4). Researchers working with Seekers are facing similar difficulties in the United States,[57] so the problem seems to be universal. Such is the flaw of unpaid voluntary collaboration. The next edition of this research, if it comes, must come with funding and paid contracts to eliminate this specific problem.

In the 2010 pilot study, intended group rotation was not achieved in any of the three trials. A scheduled performance of the larp and related tests turned out to be executable only once; its repetition with group rotation was too much to demand from voluntary teacher's cooperation. This is not to be understood that the teachers were uncooperative; they simply could not facilitate enough properly scheduled meetings before the end of the semester. Experts are right: although the rotation method may be the most reliable type of pedagogical experiment, it is very difficult to implement in a public school

daily routine.[58] This was probably the most important conclusion of the pilot studies: if the rotation method seems unexecutable and the equivalent-groups unreliable, the right course should be *one-group* experiments: no more division into control groups and experimental groups, instead favoring an experimental stage (with tests) and a control stage (with tests) for a single group.

Transition to the single-group method was not an easy decision to make, as it is considered not very reliable, unable to take into account other factors that might have influenced the outcome.[59] Among these factors is non-equal difficulty of the tested content. In the one-group procedure selected for the 2013 study, the experimental stage focused on 18th century Poland, while the control stage used a different textbook unit, early 19th century Europe. Two different book units mean different content tested with a different set of questions, which may translates into non-equivalent difficulty level.[60] This would distort the results; the experimental (larp) stage could have led to higher scores not because the method was superior but because its tests were easier. However, when the choice is between "the same edu-content with different learner groups" and "different edu-content with the same people" variables, the latter prevails.

Neither allows for 100% equivalence of the compared samples, but in the case of content/ test difficulty, equivalence can be at least partially secured by strict adherence to the book, thorough content analysis, and—first and foremost—careful design of tests: the same number, type, and difficulty of questions addressing the same thematic issues (see above under Apparatus). This is a strong advantage of the one-group method: the difficulty level of book units and test questions can be compared on the basis of recorded data and consulted with subject-matter experts. Anyone who wants to question the equivalence of compared samples can study both book units and back up his or her argument with evidence. By contrast, in the equivalent-groups method, the differences reside in humans, leaving no chance for later verification of their equivalence. Therefore, if sufficient care is given to assure equal difficulty of post-larp and non-larp tests, it should be more reliable than the equivalent-groups method.

....

56. D'Agostino, 123.

57. Sarah Lynne Bowman and Anne Standiford, "Educational Larp in the Middle School Classroom: A Mixed Method Case Study," *International Journal of Role-playing* 5. (In press for 2015 publication).

....

58. Łobocki, *Metody*, 113; Zaczyński, *Praca*, 102.

59. Maszke, *Metody*, 182.

60. D'Agostino, "Measuring," 120.

Table 6: Groups Equivalence vs. Content Equivalence

Problems with equivalence of groups (in two-groups method)	Problems with equivalence of content (in one-group method)
Once the groups are selected, the researcher can do nothing to increase equivalence	Equivalence can be increased by careful design of learning materials and tests
Differences in group competence hidden in people -- difficult to assess, impossible to verify without access to group members	Differences in content difficulty are verifiable through analysis of textbook material -- accessible for any researcher

Table 7: Suggested Data Analysis based on Individual Performance

Student number	Post-larp 1 week	Post-larp 1 month	Non-larp 1 week	Non-larp 1 month
1	10 points	8 points	10 points	9 points

A small room for non-equivalence remained in the changing number of tested samples; if the "1 week" test in Group 3 was taken by 32 students and "1 month" by 29, the sample size changed by 3. Even if the number was the same in both tests, the group was not always 100% identical.[61] The 2013 project accepted this level of attrition in order to minimize the burden put on the teachers as unpaid voluntary work. They were asked only to eliminate the tests of students who had not participated in the larp. More accurate data would have come from reports tracing the performance of individual students, e. g. Table 7, with only those who took part in the larp and submitted all four tests taken into account. If this research is continued in the future, it should adopt this stricter policy of elimination.

CONCLUSION

The idea of monodisciplinary and knowledge-oriented edu-larp worked well with the teachers, with most of them interested in further cooperation. It is yet to be seen if or when the teachers are ready to run print-and-play scenarios on their own without larpmasters coming from outside of school. It was widely accepted by students with a high engagement ratio (90%, as assessed by organizers).

When it comes to quantitatively measured educational efficiency, in all studies since 2010 with various research methods, edu-larp generally correlates with higher—or comparable, at worst—test scores than traditional revision classes, especially in long term knowledge retention. Still, the hypothesis of larp's superiority has not been substantially verified due to the unexpectedly small sample. With highly encouraging preliminary results, research should be continued, as "replication… strengthens the credibility of evidence."[62]

The evolution of procedures from pilot studies to the main study illustrate the value of constant improvement of research methods in subsequent iterations. The ultimately selected one-group method makes a strong case for its reliability as compared to alternatives and has been field-tested for its applicability in the school environment. Should other teachers and researchers be willing to follow this model, or get inspired by it, I would be very interested in the results.

....
62. Angela M. O'Donnell, "Experimental Research in Classrooms," in *Empirical Methods for Evaluating Educational Interventions*, edited by Gary D. Phye, Daniel H. Robinson and Joel Levin (San Diego: Elsevier Academic Press, 2005), 215.

....
61. Valentine and Cooper, "Can We Measure," 107-108.

Appendix A: Game Materials

Larps used in pilot studies and the main study contained three types of game materials: 1) instruction for gamemasters (available only to them), 2) one-sheet summary of the major period events players should know (one copy for everyone), 3) role cards with description of the character and his/her tasks in the game (individual). Content was based on a high-school history handbook by M. Kamiński and R. Śniegocki, covering a large unit on the Polish-Lithuanian Commonwealth in the second half of the 18th century.

a) Instruction for Gamemasters

The larps were administered by experienced larpmasters who did not need general guidelines. They received a brief description of the formula elaborated upon in the IJRP paper (Mochocki 2013, 67-68): a larp designed as a negotiation game for conflicted-yet-cooperating factions to be played in three stages:

1. Prepping: each faction meets separately, their aims being to get into character get acquainted with other faction members and their opinions; and discuss the course of action for the upcoming meeting. Playing time: 15-20 minutes.

2. Unofficial talks: all factions meet in one place before the official session and break down into small groups; trying to cut deals; form alliances; recruit followers; obtain/trade information; blackmail or bribe adversaries, etc. Playing time: about 60 minutes.

3. Official debate: formal negotiations at a table, potentially leading to a resolution, e.g. voting. Playing time: about 20 minutes.

In addition, the gamemasters received a list of characters divided into factions with suggestions for modifications to the character cast: which roles should be dropped first if the number of players is smaller and which roles are gender neutral and could be easily replaced with the opposite sex without changes in tasks or character details.

b) Unit summary

This particular scenario, called "Facing the Second Partition," was used in all games in the 2013 edition (tested in the 2010 pilot study) and was set in February 1794 near the end of the period covered by the school textbook. It was the grim history of Polish-Lithuanian Commonwealth in 18th century, when it was incrementally carved into pieces by its three neighbors. As players, students would know that the Second Partition, which meant yet another substantial loss of territory, had sparked a desperate rebellion that failed hopelessly and was punished with the Third Partition, i.e. the complete annihilation of the state. As characters, the students were faced with the Second Partition and the question of starting the rebellion was theirs to discuss.

The one-sheet unit summary had three parts:

1) Bullet-point timeline of major events. There is no room to print it here; readers interested in historical context for the below-described factions and challenges can research the Partitions of Poland online (e. g. on Wikipedia).

2) Brief characteristics of all in-game factions, based on former parties of the Great Sejm (parliament) of 1788-92:

Hetmans' Party, led by hetmans Branicki and Rzewuski, general Potocki and prince Poniński (later leaders of the Targowica Confederation):

- supporters of the old Sarmatian tradition, enemies of French novelties (especially of revolutionary ones)
- supporters of the Cardinal Laws of 1767 and of Russian supervision over the Polish-Lithuanian Commonwealth.

Patriotic Party, led by Sejm marshal Małachowski, Ignacy Potocki, Stanisław Potocki, Adam Czartoryski (ideological leader: bishop Kołłątaj, who wasn't a member of parliament)

- supporters of the enlightened reform of the state (Constitution of May 3)
- enemies of Russia; until recently, supporters of strategic alliance with Prussia
- don't like the King for his subservience to Russia

Royal Party, the King and his followers

- supporters of the enlightened reform of the state
- BUT: want to keep good relations with Russia

(while the Great Sejm was proceeding, the King allied with the Patriotic Party)

None of the influential figures listed here were played by students, but many player characters had connections to them.

3) Introduction to the opening scene of the game:

Facing the dire situation and the imminent reduction of the army, some of the patriots press for a new war. They are being contacted by repentant Targowica members, who no longer support Russia after the shameful Second Partition. The king is quietly following Russian instructions for the time being, but is also looking back at the attitude of the nobility. In great secret, members of all three former Great Sejm factions are meeting in Warsaw on trilateral talks. Things that must be decided:

- Is reconciliation between the patriots and the targowicans possible?
- Should (or when should) war with Russia be restarted (and what about Prussia)?
- Should the Constitution of May 3 be reinstalled or should Cardinal Laws remain in place?
- Should king Stanisław August Poniatowski be deposed?

c) Role cards
Each player was given an individual role card with character description and a numbered list of tasks (usually three, rarely two or four). Here is a sample:

Castellanea of Calisia

You come from the proud Greater Poland aristocracy, distantly related to the Potockis. Your family used to support the old traditions of Golden Liberty, but when the "liberty" became enforced with Russian bayonets, it was time for a change of mind. As an educated and experienced lady, you now support the enlightened thought and the Constitution of May 3. You deeply regret the collapse of the Constitution after the Russian military intervention and you feel even more hurt by the Second Partition of the country. You dream of a great and victorious war bringing independence and sovereignty back to the Commonwealth. You do understand, however, that right now the Polish forces are too weak and none of the other European countries are coming to help. It is said the patriots are preparing a new rebellion; they need to be talked back to their senses as the new war can only bring new disasters.

Task 1: Strongly advise against any military actions against Russia. At the moment, 100,000 Russian troops are stationed on Polish soil and the Prussians and Austrians cannot wait for the opportunity to grab more. War has to be postponed to a convenient time when the hostile army goes to fight somewhere else. Without a strong ally, the Commonwealth stands no chance.

Task 2: Do not let the patriots slip back into the backward Sarmatian ideology. The restitution of the Constitution of May 3 must be the primary goal, as its reforms are the only way to bring the nation back to power.

Task 3: Convince the congregation that King Poniatowski cannot be trusted, because he is likely to switch to the side of the Russians again when times get hard. But, he must be tolerated as necessary evil. After all, it is better to have him on the throne rather than some new puppet sent from Moscow.

Most tasks revolved around the initial questions listed on the summary sheet, which reflected major conflict of ideas and interests of the period: textbook-based curricular content.

Bibliography

Aarebrot, Erik and Martin Nielsen. "Prisoner for a Day: Creating a Game Without Winners." In *Playing the Learning Game: A Practical Introduction to Educational Roleplaying*, edited by Martin Eckoff Andresen, 24-29. Oslo, Norway: Fantasiforbundet, 2012.

Balzer, Myriel. "Immersion as a Prerequisite of the Didactical Potential of Role-Playing." The *International Journal of Role-Playing*, Issue 2 (2011): 32-43. http://www.ijrp.subcultures.nl/wp-content/issue2/IJRPissue2-Article3.pdf

Blatner, Adam. "Creative Drama and Role Playing in Education." In *Interactive and Improvisational Drama*, edited by Adam Blatner. New York: iUniverse, 2007. Kindle edition.

Bolton, Gavin. "A Conceptual Framework for Classroom Acting Behavior." Ph.D. diss., Durham University, 1997. http://etheses.dur.ac.uk/1627/

Bowell, Pamela and Brian S. Heap. *Planning Process Drama: Enriching Teaching and Learning*. London: Routledge, 2013. Kindle edition.

Bowman, Sarah Lynne. "Educational Live Action Role-playing Games: A Secondary Literature Review." In *Wyrd Con Companion Book 2014*, edited by Sarah Lynne Bowman. Los Angeles, CA: Wyrd Con, 2014. (In this anthology).

--. *The Functions of Role-playing Games: How Participants Create Community, Solve Problems, and Explore Identity*. Jefferson, NC: McFarland & Company, Inc., 2010.

Bowman, Sarah Lynne and Anne Standiford. "Educational Larp in the Middle School Classroom: A Mixed Method Case Study." *International Journal of Role-playing* 5 (In press for 2015 publication).

--. "Edu-larp in the Middle School Classroom: A Qualitative Case Study." In DiGRA (Digital Games Research Association), Proceedings from the RPG Summit at DiGRA 2014. Salt Lake City, UT. 3-6 August 2014. (In press for 2014 publication).

Carroll, John and David Cameron. "Playing the Game, Role Distance and Digital Performance." *Applied Theatre Researcher* 6, no. 1 (2005): 1-11. http://www.griffith.edu.au/__data/assets/pdf_file/0009/54945/playing-game.pdf

Collins, Alan, Diana Joseph and Katerine Bielaczyc. "Design Research: Theoretical and Methodological Issues." *Journal of the Learning Sciences* 13, no. 1 (2004): 15-42. http://treeves.coe.uga.edu/EDIT9990/Collins2004.pdf

Crookall, David, Rebecca Oxford and Danny Saunders. '"Towards a reconceptualization of simulation: From representation to reality." *Journal of SAGSET (Society for the Advancement of Games and Simulations in Education and Training* 17, no. 4 (1987): 147-171.

D'Agostino, Jerry. "Measuring Learning Outcomes: Reliability and Validity Issues." In *Empirical Methods for Evaluating Educational Interventions*, edited by Gary D. Phye, Daniel H. Robinson and Joel Levin, 113-146. San Diego: Elsevier Academic Press, 2005.

Echeverría, Alejandro, Christian García-Campo, Miguel Nussbaum, Francisca Gil, Marco Villalta, Matias Améstica, and Sebastian Echeverría. "A Framework for the Design and Integration of Collaborative Classroom Games." *Computers & Education* 57, no. 1 (August 2011): 1127-1136.

Galarneu, Lisa. "Authentic Learning Experiences Through Play: Games, Simulations and the Construction of Knowledge." In *Changing Views: Worlds in Play. Proceedings of DiGRA 2005 Conference*, 2005. http://www.digra.org/dl/db/06276.47486.pdf

Guscott, Steve. "Role Play as Active History." In *A Practical Guide to Teaching History in the Secondary School*, edited by Martin Hunt, 37-48. New York: Routledge, 2007.

Harder, Sanne. "Confessions of a Schoolteacher: Experiences with Role-playing in Education." In *Lifelike*, edited by Jesper Donnis, Morten Gade, and Line Thorup, 228-235. Copenhagen, Denmark: Projektgruppen KP07, 2007. http://nordiclarp.org/w/images/a/af/2007-Lifelike.pdf

Harviainen, J. Tuomas, Timo Lainema, and Eeli Saarinen. "Player-reported Impediments to Game-Based Learning." In Proceedings of DiGRA Nordic 2012 Conference: Local and Global—Games in Culture and Society, 2012. http://www.digra.org/dl/db/12168.02279.pdf

Heathcote, Dorothy. "Contexts for Active Learning—Four Models to Forge Links Between Schooling and Society." Presented at NATD Conference 2002, accessed May 10, 2012, http://www.moeplanning.co.uk/wp-content/uploads/2008/05/dh-contexts-for-active-learning.pdf

--. "Productive Tension. A Keystone in 'Mantle of the Expert' Style of Teaching." *Journal for Drama in Education* 26, no. 1 (2010): 8-23. http://www.mantleoftheexpert.com/wp-content/uploads/2010/03/Jan-101.pdf

Henriksen, Thomas Duus. "Learning by Fiction." In *As LARP Grows Up—Theory and Methods in LARP*, edited by Morten Gade, Line Thorup and Mikkel Sander, 110-115. Copenhagen: Projektgruppen KP03, 2003. http://nordiclarp.org/w/images/c/c2/2003-As.Larp.Grows.Up.pdf

--. "Moving Educational Role-Play Beyond Entertainment." *Teoria de La Educacion en la Sociedad de la Informacion* 11, no. 3 (2010): 226-262. http://es.youscribe.com/catalogue/libros/conocimientos/ciencias-humanas-y-sociales/moving-educational-role-play-beyond-entertainment-1781104/

--. "On the Transmutation of Educational Role-play: A Critical Reframing to the Role-play in Order to Meet the Educational Demands." In *Beyond Role and Play: Tools, Toys, and Theory for Harnessing the Imagination*, edited by Markus Montola and Jaakko Stenros, 107-130. Helsinki, Finland: Ropecon ry, 2004. http://nordiclarp.org/w/images/8/84/2004-Beyond.Role.and.Play.pdf

Heyward, Paul. "Emotional Engagement Through Drama: Strategies to Assist Learning through Role-play." *International Journal of Teaching and Learning in Higher Education* 22, no. 2 (2010): 197-203. http://www.isetl.org/ijtlhe/pdf/IJTLHE751.pdf

Hyltoft, Malik. "Four Reasons Why Edu-Larp Works." In *LARP: Einblicke*, edited by Karsten Dombrowski, 43-57. Braunschweig, Germany: Zauberfeder Ltd, 2010. http://www.zauberfeder-verlag.de/Produkte/MP10/MP10_Artikel-04.pdf

--. "Full-Time Edu-larpers: Experiences from Østerskov." In *Playing the Learning Game: A Practical Introduction to Educational Roleplaying*, edited by Martin Eckoff Andresen, 20-23. Oslo, Norway: Fantasiforbundet, 2012.

Jansen, Frida Sofie. "Larp the Swedish Way: An Educational Space Odyssey." In *Playing the Learning Game: A Practical Introduction to Educational Roleplaying*, edited by Martin Eckoff Andresen, 30-35. Oslo, Norway: Fantasiforbundet, 2012.

Kapp, Karl M. *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. Hoboken, NJ: John Wiley and Sons. Kindle edition, 2012.

Kot, Yaraslau I. "Educational Larp: Topics for Consideration." In *Wyrd Con Companion 2012*, edited by Aaron Vanek and Sarah Lynne Bowman, 118-127. Los Angeles, CA: Wyrd Con, 2012.

Larsson, Elge. "Participatory Education: What and Why." In *Beyond Role and Play: Tools, Toys and Theory for Harnessing the Imagination*, edited by Markus Montola and Jaakko Stenros, 243-247. Helsinki: Ropecon ry, 2004. http://nordiclarp.org/w/images/8/84/2004-Beyond.Role.and.Play.pdf

Levin, Joel. "Randomized Classroom Trials on Trial." In *Empirical Methods for Evaluating Educational Interventions*, edited by Gary D. Phye, Daniel H. Robinson, and Joel Levin, 3-27. San Diego, CA: Elsevier Academic Press, 2005.

Łobocki, Mieczysław. *Metody i techniki badań pedagogicznych*. Kraków: Oficyna Wydawnicza "Impuls," 2003.

Marks, Helen M. "Student Engagement in Instructional Activity: Patterns in the Elementary, Middle, and High School Years." *American Educational Research Journal* 37, no. 1 (Spring 2000): 153-184., 1995.

Szewkis, Eyal, Miguel Nussbaum, Fernanda Denardin, Jose P. Abalos, Tal Rosen, Daniela Caballero, Arturo Tagle and Cristian Alcoholado. "Collaboration between large groups in the classroom." *International Journal of Computer Supported Collaborative Learning* 6, no. 4 (December 2011): 561-575.

Ustawa z dnia 7 września 1991 o systemie oświaty. (Act on Education System.) http://www.dziennikustaw.gov.pl/DU/2004/s/256/2572/1/

Utne, Torstein. "Live Action Role Playing—Teaching through Gaming." In *Dissecting LARP—Collected Papers for Knutepunkt,* edited by Petter Bøckman and Ragnhild Hutchison, 23-30. Oslo: Knutepunkt, 2005. http://nordiclarp.org/w/images/9/95/2005-Dissecting.Larp.pdf

Valentine, Jeffrey C. and Harris M. Cooper. "Can We Measure the Quality of Causal Research in Education?" In *Empirical Methods for Evaluating Educational Interventions*, edited by Gary D. Phye, Daniel H. Robinson, and Joel Levin, 85-112. San Diego: Elsevier Academic Press, 2005.

Villalta, Marco, Ignacio Gajardo, Miguel Nussbaum, Juan J. Andreu, Alejandro Echeverria, and J.L. Plass. "Design Guidelines for Classroom Multiplayer Presential Games (CMPG)." *Computers & Education*, 57, no. 3 (November 2011): 2039-2053.

Zaczyński, Władysław. Praca badawcza nauczyciela. Warszawa: Wydawnictwa Szkolne i Pedagogiczne, 1995.

Michał Mochocki, *Ph.D. in Literature, works as Assistant Professor at Kazimierz Wielki University in Bydgoszcz, Poland. He teaches literature and culture at English Studies and is the leading force behind a newly created GAMEDEC: Game Studies & Design BA Programme. Mochocki co-authored the Dzikie Pola tabletop RPG and worked for several years as editor and author of RPG content for Polish RPG magazines. He has been a larper and larp designer since 2000, specializing in historical larps and edu-larps, sometimes combined with historical reenactment. Mochocki works as edutainment consultant with Nowa Era educational publisher. Now, at Gamedec. UKW, he teaches a number of courses in non-digital game design, most notably tabletop RPGs and LARP. Websites: michal-mochocki.pl, mmochocki.blogspot.com, gamedec.ukw.edu.pl.*