Revisions Made to Manuscript

The authors would like to thank the referees for their helpful comments and suggestions. The revisions made and other comments are listed below.

No	Reviewer Comment	Response/Action	Reference to where change is made
		Reviewer A	
		Abstract	
1	…amongst others, to generate…	Changed as suggested.	Page 1, Abstract, line 3.
2	"generalized solution" works if we use the definition of solution as "A means of solving a problem.", and your meaning is clear here, but please check throughout the paper that you are not using solution as "A correct answer", were "generalized solution" could be confused as meaning a particular solution that satisfies several problems. Perhaps consider using another phrase to avoid the ambiguity.	A footnote explaining the difference has been added.	Page 1, Introduction, line 11.
3	Hyper-heuristics, on the other hand, search a heuristic space ~~instead~~ with…	Changed as suggested.	Page 1, Abstract , line 5.
4	…new techniques ~~which~~ that has…	Changed as suggested.	Page 1, Abstract, line 7.
5	A two-phased approach is taken, with the first phase focusing on hard constraints, and the second…	Changed as suggested.	Page 1, Abstract, lines 10 to 11.
6	…choose parents, to which the mutation and crossover… are applied.	Changed as suggested.	Page 1, Abstract, lines 12 to 13.
7	…problems, thereby providing…	Changed as suggested.	Page 1, Abstract , line 17.
		Introduction (P1)	
8	…algorithms, amongst others [28].	Changed as suggested.	Page 1, Introduction, line 8.
9	…~~the~~ travelling salesman ~~problem~~, and vehicle routing problems, and ~~has~~ have…	Changed as suggested.	Page 1, Introduction, line 15.

10	…the most effective mutation operators ~~in solving this problem was~~ were problem dependent,…	Changed as suggested.	Page 1, Introduction, line 17.
11	…generalizing over a set of problems, with different sets…	Changed as suggested.	Page 1, Introduction, line 20.
	Introduction (P2)		
12	…five different school timetabling problems, and was able to produce feasible solutions of good quality for each problem~~.the five different types of school timetabling problems.~~	Changed as suggested.	Page 2, Introduction, paragraph 2, lines 5 to 6.
13	…terms of school timetabling, ~~hyper-heuristics,~~ evolutionary algorithms, ~~and~~ hyper-heuristics, and…	Changed as suggested.	Page 2, Introduction, paragraph 2, line 8.
14	…Section 6.	Changed as suggested.	Page 2, Introduction, paragraph 2, line 12.
	Background (P2)		
15	…problem, and evolutionary…	Changed as suggested.	Page 2, Background, paragraph 1, line 2.
	2.1. (P2)		
16	…timetable periods, so as to satisfy the problem hard constraints, and minimize	Changed as suggested.	Page 2, Section 2.1, paragraph 1, lines 1 to 2.
17	In some problems, venues are…	Changed as suggested.	Page 2, Section 2.1, paragraph 1, line 3.
18	…minimized, and…	Changed as suggested.	Page 2, Section 2.1, paragraph 1, line 5.
19	…next, and what may be defined as a hard constraint for one problem could be regarded as a soft constraint for another, and vice versa.	Changed as suggested.	Page 2, Section 2.1, paragraph 2, lines 1 to 3.
	2.1. (P3)		
20	…jump down algorithm, as well as various hybrid approaches…	Changed as suggested.	Page 3, Section 2.1, paragraph 1, line 16.
21	…by means of two processes, namely heuristic selection…	Changed as suggested.	Page 3, Section 2.2, paragraph 2, line 9.
22	Please check the rest of the paper for any similar comma problems.	Checked and changed as suggested.	Throughout paper.

	2.2 (P3)		
23	Deterministic move acceptance methods either accept all moves or select moves resulting in equal and/or improved cost; while methods for non-deterministic acceptance include Monte Carlo,	Changed as suggested.	Page 3, Section 2.2, paragraph 2, lines 15 to 16.
	2.2 (P4)		
24	Given the progress made with selection perturbative hyper-heuristics and their effectiveness…	Changed as suggested.	Page 4, paragraph 3, line 1.
	2.4 (P5)		
24	Both GAs and GP first create an initial population of randomly created chromosomes or individuals, which are then iteratively refined…	Changed as suggested.	Page 5, Section 2.4, paragraph 1, lines 2 to 3.
	3.1.1 (P6)		
25	One violation row ~~some~~ swap	Changed as suggested.	Page 6, Section 3.1.1, 5th bullet
26	"It is anticipated that one or a low number of swaps will improve the individual but possibly not at a sufficient rate and could possibly lead to long runtimes or an unsuccessful hyper-heuristic." **– I would imagine this depends entirely on the algorithmic implementation. A high number of swaps could also lead to degradation of the solution quality. This is something that should be tested rather than surmised upon. Perhaps leave this out.**	Removed as suggested.	Page 6, Section 3.1.1, first paragraph after bullets.
	3.1.2 (P7)		
27	Here *n* and *m* are used to denote different variables than appear in previous sections. Never give a variable name multiple meanings in a single body of work.	The variable representing the number of tuples in Section 2.1 has been changed from *n* to *x* in Section 2.1. The variable representing the number of generations in Section 3 has been changed from *m* to *g*. The variable representing the number of swaps in section 3.1.1 has been changed from *n s* swaps. Variables *n* and *m* are only used to represent the number of chromosomes created by the sequential construction method and the GA population size respectively.	Page 3, paragraph 1, line 6. Page 5, Section 3, line 9. Page 6, 1st paragraph after bullets, line 1. Page 8, Section 3.1, paragraph 1, line 1 and paragraph 2, line 4.

28	…created by firstly creating n chromosomes. The chromosome amongst these with…	Paragraph has been rewritten for clarity and this has been changed.	Page 8, Section 3.1.2, paragraph 1, lines 1 to 2.
29	… violations, then forms…	Paragraph has been rewritten for clarity and this has been removed.	N/A
30	Paragraph 1: Should mention that the chromosome elements are selected at random.	Included.	Page 8, Section 3.1.2, paragraph 1, lines 1 to 2.
31	I am confused as to the relationship between the chromosomes and the timetable during Phase 1. There is the sentence "The chromosome is used to create a timetable.", followed by a timetable construction method that makes no reference to the chromosome. How is the chromosome applied to the timetable? Is each operation applied only once from left to right? Is the full chromosome applied more than once? Please resolve this confusion. A formal pseudo code algorithmic formulation would be very helpful in Section 3.	The section has been rewritten for clarity and the pseudo code included.	Page 8, Section 3.1.2 including Figure 3.
	4 (P8)		
32	Please number the objective functions.	Numbered as suggested.	Page 10.
33	In both functions, *Min* should not be in italics – could be confused as variables.	Changed as suggested.	Page 10, equation (1) and (2).
	4 (P9)		
34	… (hence the acronym hdtt)…	Changed as suggested.	Page 10, Section 4.1, paragraph 1, line 2.
	4.1-4.4 (P9-11)		
35	Please distinguish between the constraint labels of different problems – e.g. HC^1_1 and HC^2_1 for problems 4.1 and 4.2.	Changed as suggested.	Pages 10 to 12, sections 4.1 to 4.4.
	5 (P11)		
36	Subpopulation size (n) – During Phase 1 each element of the population is created by firstly …	Changed as suggested.	Page 12, Section 5, second bullet.

37	I am surprised at this high number of swaps per mutation operation – mutation usually occurs with a low probability and has a small effect on the chromosome, otherwise it may easily result in genetic drift, and the algorithm becomes little better than a random heuristic selector. Perhaps I am misunderstanding what "number of swaps" entails. Please make it clearer how this corresponds to the discussion of mutation in section 3.3.	This refers to the number of swaps performed by the mutational low-level perturbative heuristics, not the mutation operator of the genetic algorithm. This has been clarified in the manuscript.	Page 12, Section 5, 5th bullet. Page 13, Table 3.
		6 (P12)	
38	The algorithmic comparison here is probably the best one can do given the limited information and time constraints, and shows a definite qualitative improvement in the results produced by GASPHH. One of the biggest problems I encounter with comparison studies between authors is that different convergence / runtime criteria are used (sometimes using an arbitrary number of generations) and typically using different programming methods and computing platforms, making even a casual comparison of algorithmic performance problematic. Ideally, the researcher should implement the compared methods themselves to level the playing field, but this is not always practical and should not be necessary. The operations research community must develop and follow standards for algorithm comparison testing.	The authors have taken an example from previous studies (e.g. 7, 25, 35 listed in the references) in this domain and performed an empirical comparison of results for different methods applied to the same problem instances. It is agreed that a better methodology for comparison of algorithm performance needs to be identified and this will examined as part of future work.	N/A
		7 (P15)	
39	…evaluated over ~~more than one~~ multiple generations.	Changed as suggested.	Page 17, Section 7, line 13.
		Reviewer B	
		Major issues	
1	The authors should present the exact equation of the fitness function they use and explain it in detail.	The same equation is used for all problem instances. The equations for the fitness function for both phases and the explanation thereof is presented in section 4.	Page 10, Section 4, equations (1) and (2).

2	The authors should present the pseudo code or flowcharts of the proposed algorithm in order to be easy for the reader to understand it.	Pseudo code for the genetic algorithm, application of the low-level mutational perturbative heuristics and timetable creation have been included.	Page 6, Figure 1. Page 7, Figure 2. Page 8, Figure 3.
3	The authors compare their algorithm with another one which is not published yet (it is under review). This is not acceptable. How is it possible to check the results presented?	The paper has since been published and the reference has been updated in the manuscript.	Page 19, reference 13.
4	The authors compare their algorithm with the algorithm presented by Beligiannis et al. However, only 6 instances are reported. Instance 6 is missing.	Subsequent to making these problem sets publicly available it was found that there was an error in instance 6 and it has not been used in subsequent studies (e. g. 36 and 39 in the references). The number of the problem instance has been changed to 7 (previously was listed as the sixth of 6 problems from the Beligiannis problem set).	4.2, Page 11, 1st paragraph after the 1st set of bullets, lines 2-3.
5	The execution times of the presented algorithm should be presented.	The runtimes are presented in section 6.	6, Page 14, Table 5.
6	In Table 8, the results reported concerning the algorithm of Beligiannis et al. are different from the ones reported in the original paper. Why is this so?	The study presented in the paper evaluates timetables for all four soft constraints listed on page 1267 of the original paper, however the results reported in the original paper (Table 1 - lists the cost of each soft constraint in a separate column) does not include the uniform distribution of idle periods for teachers constraint. The authors have written to Beligiannis et al. to obtain the best timetables produced by the method and tested these for all four constraints. These are the values listed in Table 8 of this paper.	N/A

7	The authors should refer to more recent papers that use these instances like: a. A simulated Annealing with a new neighborhood structure based algorithm for high school timetabling problems, European Journal of Operational Research, by D. Zhang, Y. Liu et.al. b. A hybrid particle swarm optimization based algorithm for high school timetabling problems, by I.X. Tassopoulos, G.N. Beligiannis, Applied Soft Computing.	Reference to both these papers have now been made in section 2.1 and a comparison of the results obtained by Zhang et al. for the Abramson problem set has been included in Table 6. However, a comparison with the results presented in these papers for the Beligiannis data set could not be conducted as the fourth constraint, namely, the uniform distribution of idle periods for teachers is not included (see comment 6 above). The best timetables found are also not publicly available so it is not possible to apply the evaluator used in this study to these and report the results for comparison purposes.	Page 3, Section 2.1, paragraph 1, line 17. Page 15, Table 6.
8	The authors present their results on only one of the three instances presented in Valouxis et al. They should apply their algorithm to the other two instances, too.	In the paper by Valouxis et al., only one data set instance is made available as well as a sample solution. The other two instances were not available and attempts to contact the authors to obtain these were unsuccessful.	N/A
		Minor issues	
1	Page 4, Table 8: the titles of the columns are wrong.	The column headings correspond to the problem instance, evolutionary algorithm implemented by Beligiannis [7], the GA [31] and GASPHH. The values listed in the last three columns correspond to the total soft constraint cost of the best timetable produced by each of the methods. This is explained in the paper.	N/A